Cysteine-Free Native Peptide Ligation for the Assembly of Glycoproteins

David W. Anderson

PhD

University of Edinburgh

2006

Contents

Declaration v

Abstract vi

Abbreviations vii

I. Introduction

1. Biological importance of glycoproteins 1

1.1 *N*-linked glycosylation 4

1.2 *O*-linked glycosylation 9

2. Assembly of glycoproteins and glycoprotein mimics 14

2.1 Glycoproteins from biological sources 15

2.2 Synthesis of glycopeptides and glycopeptide mimics 19

2.2.1 Peptide synthesis 19

2.2.2 Oligosaccharide synthesis 21

2.2.3 Formation of the oligosaccharide-peptide linkage 26

2.3 Assembly of glycoproteins 38

2.3.1 Native chemical ligation 38

2.3.2 Expressed protein ligation 45

2.3.3 Cysteine-free peptide ligation 48

2.3.4 Post ligation modification of cysteine and selenocysteine 49

2.3.5 Acyl transfer auxiliaries 51

2.3.6 "Traceless" Staudinger ligation 57

2.3.7 Peptide ligation via decarboxylative condensation 62

2.3.8 *C*-Terminal activation methods 64

2.3.9 Protease-catalysed peptide ligation 67

2.3.10 Synthesis of thioester peptides 68

3. Project strategy 80

II. Results and Discussion

1. Synthesis of auxiliary-peptides 87

1.1 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary 88

1.1.1 Synthesis of thioethers 90

1.1.2 Synthesis of thioethers via Buchwald/Hartwig coupling 91

1.1.3 Synthesis of auxiliary-amino acid cassettes 99

1.1.4 Synthesis of auxiliary-peptides 104

1.2 *Nα*-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary 111

1.2.1 Synthesis of auxiliary-amino acid cassettes and amines 112

1.2.2 Synthesis of auxiliary-peptides 123

2. Synthesis of thioester peptides 128

2.1 Sulfonamide safety catch peptides: initial coupling with DIPCDI 128

2.2 Sulfonamide safety catch peptides: initial coupling with acid fluoride 130

2.3 "Double-linker" sulfonamide safety catch peptides 132

3. Ligation studies 146

3.1 Ligation at Gly-Gly junctions 146

3.1.1 Ligation results 146

3.1.2 Auxiliary removal 149

3.1.3 Glycopeptide ligation 151

3.2 Ligation at junctions other than Gly-Gly 155

3.2.1 Initial studies 155

3.2.2 Thiol-free ligations 158

3.3 Ligation summary 168

III. Conclusions

1. Conclusions 171

2. Future work 174

Experimental details 177

Acknowledgements 217

References 218

Appendix: Macmillan, D.; Anderson, D. W. *Org. Lett.* **2004**, *6*, 4659-4662.

(reprinted with permission, copyright 2004 American Chemical Society)

Declaration

I hereby declare that this thesis has been composed entirely by me and that all the work within it, except where clearly stated, is my own and has not been submitted for any other degree or qualification.

David Anderson

Abstract

Homogeneous, structurally defined glycoproteins can be assembled by coupling synthetic glycopeptides to synthetic or bacterially-derived protein fragments using the native chemical ligation (NCL) reaction. A limitation of NCL is the requirement for an *N*-terminal cysteine residue in one of the peptide fragments. One method for cysteine-free peptide ligation utilizes thiol acyl transfer auxiliaries which effect ligation and can then be removed under mild conditions.

We developed new, rapid routes to 1-phenyl-2-mercaptoethyl and 2-mercaptobenzyl auxiliaries. The key steps involved: 1) introduction of a suitably protected thiol to auxiliary precursors; 2) direct reductive amination of the auxiliary aldehyde or ketone to afford the auxiliary-amine, which can be conjugated to a peptide via the "sub-monomer" approach, or the auxiliary-amino acid "cassette" for use in conventional solid phase peptide synthesis. Overall yields are 53-83 %. A glycopeptide was then assembled via auxiliary-mediated ligation at a Gly-Gly junction, which was complete within 48 hours.

Thioester- and auxiliary-peptides were assembled to investigate the scope and limitations of auxiliary-mediated ligation for non Gly-Gly junctions. The 1-phenyl-2-mercaptoethyl auxiliary effected ligation at Ala-Gly, Lys-Gly and Gly-Ala junctions in 24-70 % yield, whereas the 2-mercaptobenzyl auxiliary effected ligation at a Gly-Ala junction in 42 % yield. Excess thiol was found to inhibit ligation, indicating a change in rate-determining-step relative to cysteine ligation.

Abbreviations

BINAP:	1,1-binaphthalene-2,2-diylbis-(diphenylphosphine)
Boc:	*t*-butoxycarbonyl
DBU:	1,8-diazabicylco[5.4.0]undecane-7
Ddz:	*N*-2-(3,5-dimethoxyphenyl)propyl[2]oxycarbonyl
DEAD:	diethylazodicarboxylate
DIPCDI:	diisopropylcarbodiimide
DiPPF:	1,1-bis-(diisopropylphosphino)-ferrocene
DMF:	*N*,*N*-dimethylformamide
DMSO:	dimethylsulfoxide
DPPA:	diphenylphosphoryl azide
DTT:	dithiothreitol
EDT:	ethanedithiol
EGF:	epidermal growth factor
EPL:	expressed protein ligation
EPO:	erythropoietin
ER:	endoplasmic reticulum
ERAD:	endoplasmic reticulum associated degradation
ESI:	electrospray ionisation
Fmoc:	*N*-9-fluorenylmethoxycarbonyl
GlyCAM-1:	glycosylation dependent cellular adhesion molecule-1
GPI:	glycosylphosphatidylinositol
HBTU:	2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium-hexafluoro-phosphate

HOBt:	1-hydroxybenzotriazole
HOOBt:	3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine
HPLC:	high performance liquid chromatography
LC-MS:	liquid chromatography – mass spectrometry
MESNa:	sodium 2-mercaptoethylsulfonate
MPAA:	4-mercaptophenylacetic acid/4-(carboxymethyl)thiophenol
NCL:	native chemical ligation
NMI:	1-methyl imidazole
NMP:	*N*-methylpyrrolidinone
ONB:	*ortho*-nitro benzyl
PBS:	phosphate buffered saline
PMB:	*para*-methoxy benzyl
PPTS:	pyridinium *para*-toluenesulfonate
PSGL-1:	P-selectin glycoprotein ligand-1
PyBOP:	benzotriazole-1-yl-oxy-tris-pyrrolidinophosphonium-hexafluoro-phosphate
SPPS:	solid phase peptide synthesis
TBAF:	tetrabutylammonium fluoride
TCEP:	tricarboxyethylphosphine
TFA:	trifluoroacetic acid
TFE:	trifluoroethanol
Thz:	1,3-thiazolidine-4-carboxo
TIC:	total ion count
TIPS:	triisopropylsilyl

TLC:	thin layer chromatography
tRNA:	*transfer* ribonucleic acid
UDP:	uridine diphosphate

Ala/A:	alanine
Arg/R:	arginine
Asn/N:	asparagine
Asp/D:	aspartic acid
Cys/C:	cysteine
Gln/Q:	glutamine
Glu/E:	glutamic acid
Gly/G:	glycine
His/H:	histidine
Ile/I	isoleucine
Leu/L:	leucine
Lys/K:	lysine
Met/M:	methionine
Phe/F:	phenylalanine
Pro/P:	proline
Ser/S:	serine
Thr/T:	threonine
Try/W:	tryptophan
Tyr/Y:	tyrosine
Val/V:	valine

Glc:	glucose
GlcNAc:	2-*N*-acetlyglucosamine
Gal:	galatose
GalNAc:	2-*N*-acetlygalactosamine
Man:	mannose
Fuc:	fucose

I. Introduction

1. Biological importance of glycoproteins

Protein glycosylation, the modification of the polypeptide backbone with saccharide units, (see **Figure 1**) is one of the most common and important post-translational modifications undergone by proteins and is vital for the normal growth and development of organisms. The oligosaccharide groups participate in a large number of important biological processes ranging from protein folding and secretion to cell-cell recognition and adhesion, and are therefore of fundamental importance in a diverse range of processes such as bacterial and viral infection, inflammation and cancer.[1-4]

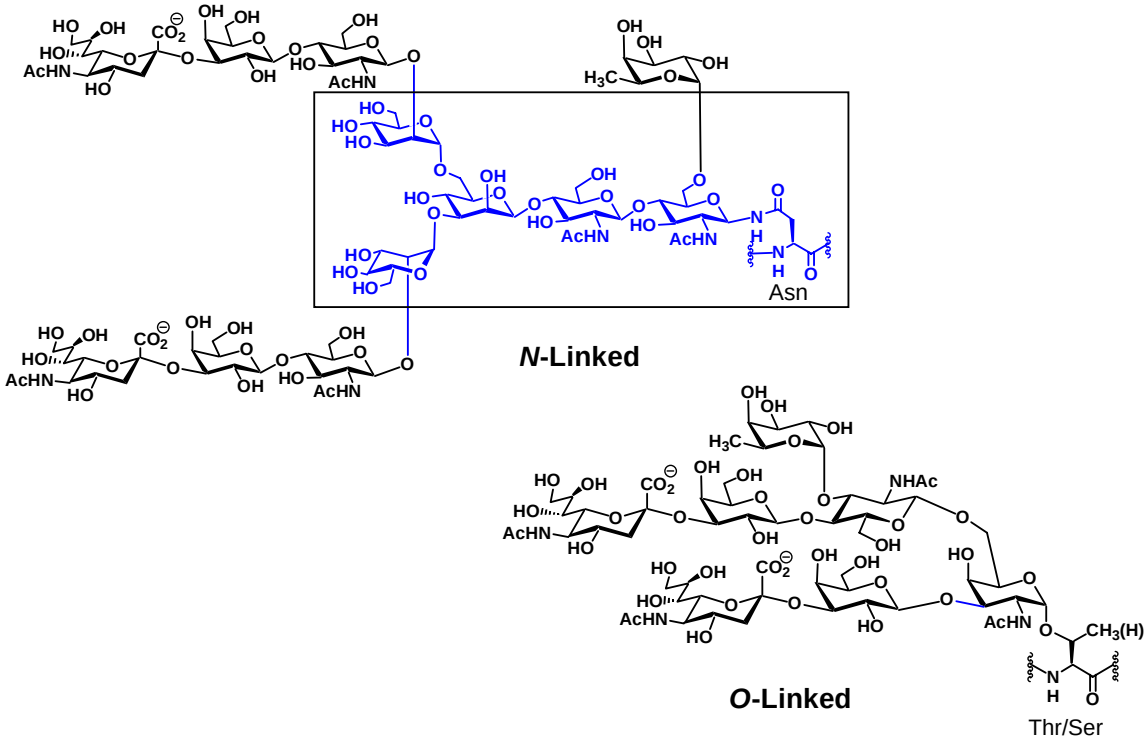


Figure 1: typical *N*- and *O*-linked oligosaccharides found on mammalian glycoproteins.

In cell-cell communication, saccharides normally have only weak binding affinity for their receptor proteins, known as lectins. When multiple oligosaccharides are displayed on a peptide scaffold in glycoproteins however, multiple receptor-oligosaccharide interactions occur, resulting in strong overall affinity of glycoproteins for their target receptors. This type of binding is highly specific, as the correct oligosaccharides need to be displayed in the correct arrangement for binding to occur.[5,6] Making only small changes in the displayed oligosaccharides or their spacing can therefore have a large effect on the binding specificity.

The oligosaccharide groups also modulate key biophysical properties of glycoproteins by forming hydrogen-bonding and hydrophobic interactions with the protein and by shielding the polypeptide backbone with a coating of saccharide groups. They therefore play crucial roles in maintaining the correct folding of glycoproteins. Indeed, even when not directly involved in binding and adhesion events as displayed ligands, the oligosaccharide groups are often vital for the protein to adopt the conformation required for binding. The oligosaccharide groups also increase protein solubility, maintain thermal stability, and shield the polypeptide backbone from proteolysis.[7,8]

Understanding of the significance of the specific oligosaccharide groups displayed by glycoproteins in cell-cell communication, and also as modulators of the biophysical properties of proteins, has been hindered by glycoprotein "microheterogeneity".[9] This phenomenon is caused by the fact that the oligosaccharide, or "glycan", groups of glycoproteins are built up in the endoplasmic reticulum (ER) and the golgi complex in a stepwise process which is not template-driven, nor under direct genetic control.[3] As a

result, each glycoprotein is expressed as a complex mixture of "glycoforms" with highly diverse glycan structures. Separating specific glycoproteins for biological study from these heterogeneous mixtures of closely-related molecules is extremely difficult and often yields only very small amounts of material.[10] Consequently, the synthesis of homogeneously glycosylated glycoproteins has been the focus of much research effort over the past decades.

The oligosaccharide, or glycan, groups of glycoproteins can be divided into two main classes based on the nature of the covalent linkage to the polypeptide backbone. *O*-linked glycans are attached via an *O*-glycosidic linkage through the alcohol side-chain of the serine or threonine amino acid residues, whereas *N*-linked glycoproteins are attached through the amide side-chain of asparagine (see **Figure 1**).

More unusual glycosidic linkages are also known. *C*-glycosides, where the C2 position of the indole side-chain of tryptophan is linked to mannose, and *S*-glycosides, where the thiol side-chain of cysteine is linked to glucose or galactose, have both been isolated from human urine, for example.[11]

1.1 *N*-linked glycosylation

N-linked glycosylation occurs in a diverse range of organisms, from yeasts and fungi to mammals.[12] The oligosaccharide group is transferred to the protein, and subsequently modified, by numerous enzymes in the endoplasmic reticulum and golgi apparatus. Modification of asparagine with an oligosaccharide group can occur for any protein which contains the amino acid consensus sequence AsnXaaSer/Thr, where Xaa represents any amino acid other than proline.

All *N*-linked glycans contain the basic core pentasaccharide structure $Man_3GlcNAc_2$(β-*N*)Asn (see **Figure 2**). This core structure is elaborated enzymatically to create a huge number of different possible glycan structures.[3,13]

N-glycans can be divided into three sub-classes based on the types of subsequent elaboration which the pentasaccharide core undergoes. Complex type *N*-glycans are typified by the absence of mannose beyond the pentasaccharide core and are extended by the addition of *N*-acetylglucosamine and galactose monosaccharides. High-mannose type glycans are extended by the addition of α-linked mannose residues, and hybrid type glycans contain elements of both complex and high-mannose types (see **Figure 2**).[8]

N-linked glycans have many diverse functions. Intracellularly, they are involved in the regulation of protein folding and the degradation of misfolded proteins. Extracellularly, *N*-linked glycans are crucial for the activity of many important cell-surface or secreted

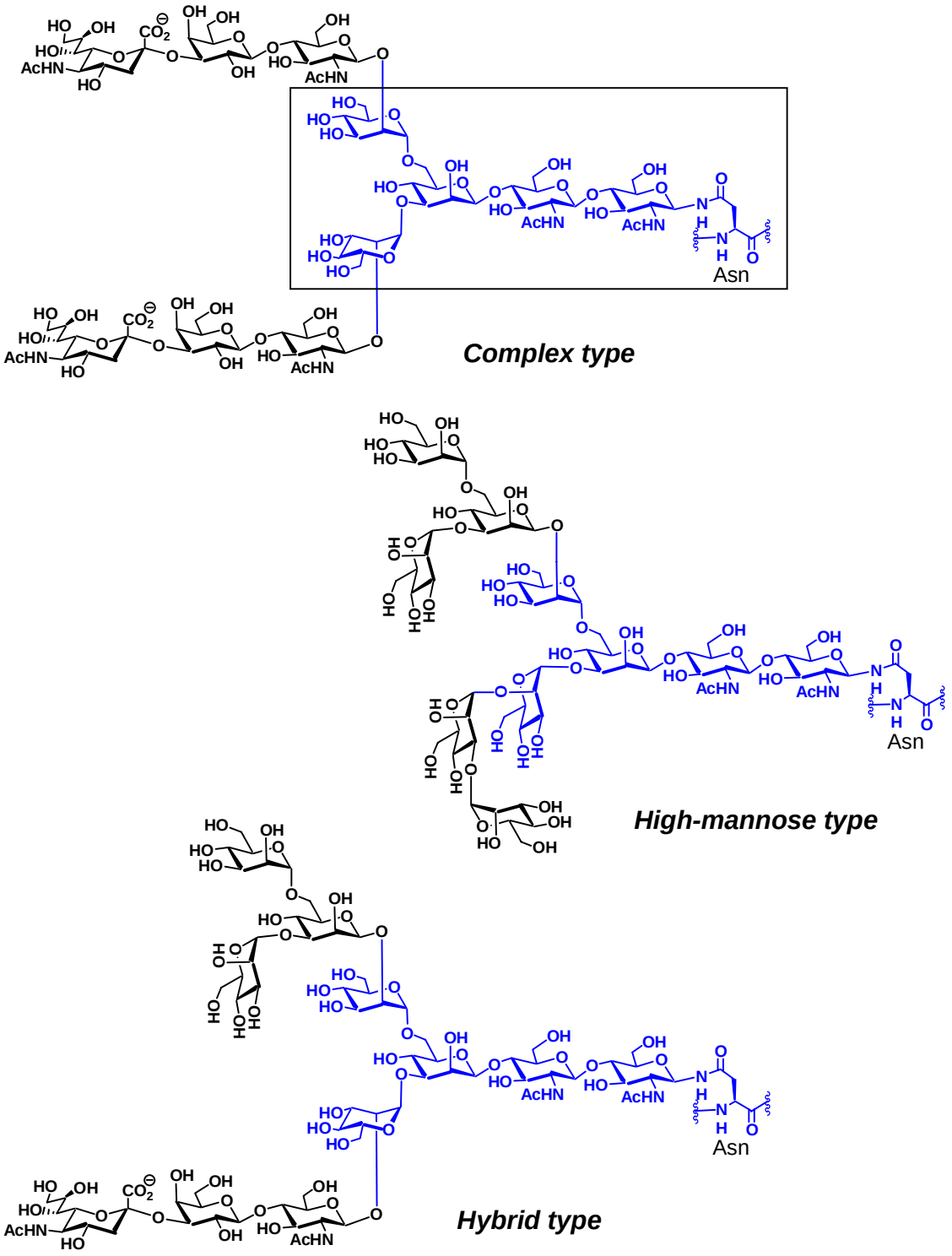


Figure 2: the three main sub-classes of *N*-linked glycans, with the core pentasaccharide highlighted.[8]

glycoproteins, either through modulation of protein structure and stability or by acting as ligands for a wide range of receptors.

The glycosylphosphatidylinositol (GPI)-anchored glycopeptide CD52 is expressed by human lymphocyte cells and is involved in the immune system. The peptide contains a single *N*-linked glycosylation site, which displays complex type *N*-glycans, which have been found to be necessary for biological activity, for example by acting as ligands for sperm-specific antibodies.[14,15]

The glycoprotein hormone erythropoietin (EPO),[16] a blockbuster treatment for anaemia which stimulates the production of red blood cells and which also exhibits neuroprotective activity, contains three *N*-glycosylation sites and a single *O*-glycosylation site.[12] The glycans make up 40 % of the mass of the glycoprotein and are vital to its *in vivo* stability and activity. EPO lacking fully elaborated *N*-glycans is rapidly cleared from the body and so has essentially no activity.[17] The terminal saccharides displayed by the oligosaccharide groups are also important. The removal of the terminal sialic acid groups from the *N*-glycans abolishes erythropoietic activity but the resulting asialoEPO retains potent neuroprotective activity in a wide range of diseases.[8,12] *N*-glycans are also utilised by a number of viruses to bind to and invade host cells. The HIV-1, hepatitis C and ebola viruses, among others, bind to host cells through high-mannose type *N*-glycans displayed on their envelope glycoproteins.[18]

The biosynthesis of *N*-linked glycans occurs in the ER and golgi apparatus. A heptasaccharide is initially assembled by cytosolic enzymes on a dolicholpyrophosphate

carrier lipid which is bound to the ER membrane. This is then "flipped" to the lumenal face of the membrane, whereupon it is further elaborated into a 14-residue oligosaccharide by ER enzymes. The oligosaccharide is then transferred to a growing protein chain at the side chain of an asparagine residue by the action of oligosaccharyltransferase, a multi subunit enzyme complex also bound to the ER membrane, which recognises the glycosylation sequence AsnXaaSer/Thr. Two terminal glucose units are subsequently trimmed off by ER-resident glucosidase enzymes to give a high-mannose type oligosaccharide ($Glc_1Man_{9\text{-}6}GlcNAc_2$) with a single terminal glucose residue (see **Figure 3**).[3,13]

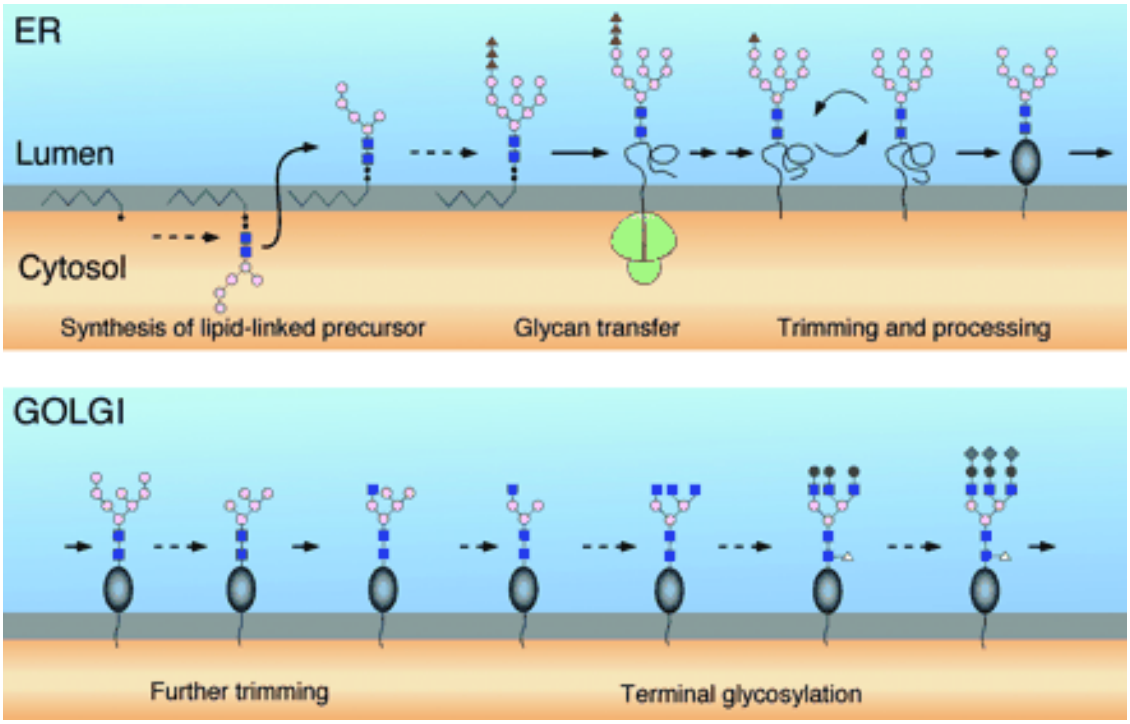


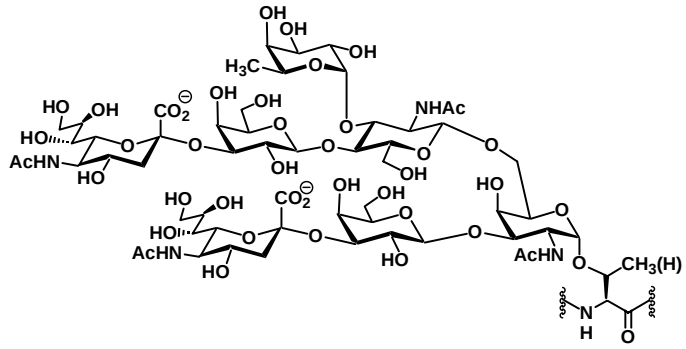
Figure 3: biosynthesis of *N*-linked oligosaccharides. ■: GlcNAc; ●: Man; ▲: Glc; ▲: Fuc; ●, ♦: variable saccharides.[3] Reprinted with permission from: Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364-2369. Copyright 2001 AAAS.

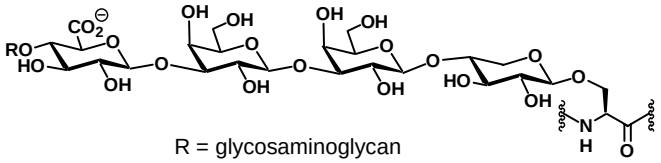
One of the most important actions of *N*-glycans is in ensuring the correct folding of the protein. This is accomplished by the calnexin-calreticulin cycle. Calnexin is bound to the membrane of the ER, whereas calreticulin is a soluble protein.[3] These complementary ER lectins bind to glycoproteins through the remaining terminal glucose of the $Glc_1Man_{9\text{-}6}GlcNAc_2$ oligosaccharide and promote correct folding. Calnexin and calreticulin also form complexes with another folding factor protein, ERp57. This thiol oxoreductase forms disulfide bonds with cysteine residues in the glycoprotein which assist folding.[19] The duration of glycoprotein to lectin binding can range from minutes to hours, and is dependent on how quickly the protein attains its native folded conformation.[3] The glycoprotein is released from calnexin and calreticulin by the removal of the terminal glucose residue from the *N*-glycan by the glucosidase II enzyme. Correctly folded proteins are then released from the ER. Misfolded proteins are recognised by the ER enzyme uridine diphosphate (UDP)-glucose:glycoprotein glucosyltransferase, which binds to the high mannose type *N*-glycan and recognises specific properties of incorrectly folded proteins such as exposed hydrophobic regions.[3,19,20] The enzyme adds the terminal glucose back to these proteins, allowing them to bind again to calnexin or calreticulin. (see **Figure 3**)

Proteins which cannot be correctly folded are retained in the ER and destroyed by endoplasmic reticulum associated degradation (ERAD). This process is not fully understood, as it involves transport to the cytosol by several pathways, including trimming by ER-mannosidase I and by interaction with calnexin. Once the correctly folded glycoprotein is released from the ER, it enters the golgi apparatus, where further trimming of the oligosaccharide occurs, before further monosaccharide units are added

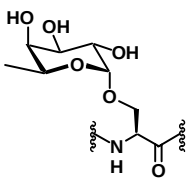
by golgi enzymes to build up the *N*-glycan in a stepwise manner, resulting in glycoprotein microheterogeneity.[3,13,19]

1.2 *O*-linked glycosylation

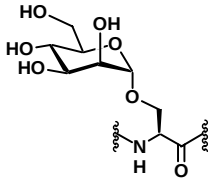




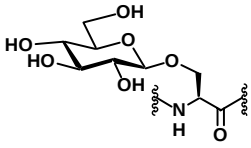
β-*O*-xylose (Proteoglycans)



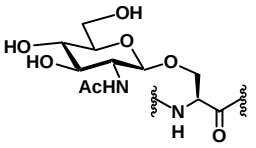
α-*O*-fucose



α-*O*-mannose



β-*O*-glucose



β-*O*-GlcNAc

Figure 4: some examples of types of *O*-linked glycans.

The modification of a hydroxyl group on the polypeptide backbone of a protein with a saccharide or oligosaccharide group occurs in a wide variety of proteins. *O*-linked glycosylation also occurs via a diverse array of linkages involving different saccharides, amino acids and stereochemistries (see **Figure 4**).

The most common form of *O*-linked glycosylation in eukaryotes is mucin type glycosylation, characterised by the attachment of *N*-acetylgalactosamine (GalNAc) via an α-glycosidic linkage to the hydroxyl side chain of serine or threonine. Mucin type glycoproteins display dense clusters of oligosaccharide groups attached via this basic structure, and hence have highly ordered extended protein conformations. Mucin type glycans can also occur at isolated sites in non-mucin glycoproteins, however.[8,12]

Proteoglycans, found in the extracellular matrix, are *O*-glycosylated with glycosaminoglycans such as heparan sulfate through β-linked xylose residues. Intracellular cytosolic and nuclear proteins contain serine and threonine residues modified with β-*O*-*N*-acetylglucosamine (GlcNAc) monosaccharides which are not further substitued by other sugar residues, in contrast to most other peptide-linked monosaccharides. α-*O*-Linked fucose residues and β-*O*-linked glucose residues are found on epidermal growth factor (EGF)-like sequences on the transmembrane protein Notch.[4,8,21] α-*O*-Linked mannose residues are found in yeast cell wall proteins and also in glycoproteins secreted by bacteria.[4] Amino acid residues other than serine or threonine can also be glycosylated. For example, both vertebrate and invertebrate collagens contain galactose residues β-linked to the side chain of hydroxylysine.[4,8]

In contrast to *N*-linked glycosylation, mucin type *O*-linked glycosylation is initiated in the Golgi apparatus with the transfer of a single GalNAc monosaccharide unit to the side chain of a serine or threonine residue. Also unlike *N*-linked glycosylation, there is no consensus sequence for the glycosylation and any serine or threonine residue is a potential site for modification, although glycosylation of serine or threonine in some sequences is unfavourable.

The GalNAc monosaccharide is added by the family of polypeptide *N*-acetyl-α-galactosaminyltransferase (ppGalNAcT) enzymes. This large family of retaining glycosyltransferases use the nucleotide donor UDP-GalNAc to modify proteins and appear to be located throughout the Golgi apparatus.[23] The properties of proteins which are recognised by ppGalNacTs are not well characterised, but appear to include β-turn protein conformations and proline-rich domains. The level of glycosylation also affects the ability of ppGalNAcTs to modify protein substrates, with some members of the family showing specificity for glycosylated proteins, and thus accounting for the prevalence of densely glycosylated mucin domains.[22,23] The X-ray crystal structure of murine ppGalNAcT-1 has recently been reported,[24] giving insights into the mechanism of the ppGalNAcT family.

The basic GalNAc(α-*O*)Ser/Thr "T_N antigen" unit is modified in the Golgi apparatus by a diverse series of downstream glycosyltransferase enzymes which results in the generation of a variety of core glycan structures depending on which specific enzymes interact with the glycoprotein (see **Figure 5**). These core structures are then elaborated further by the action of other glycosyltransferase enzymes in the Golgi apparatus to

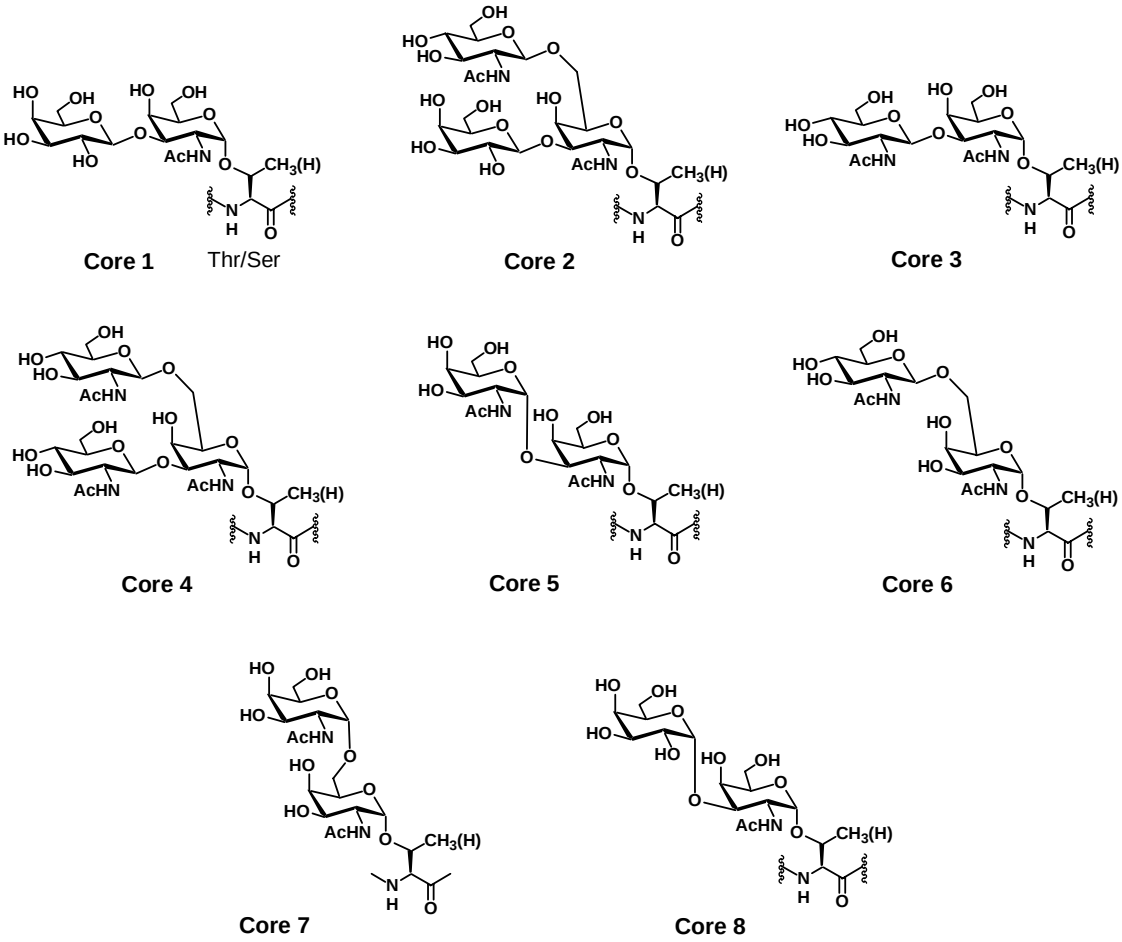


Figure 5: core *O*-linked glycans.

create diverse complex *O*-glycans.[23] The expression of the key downstream glycosyltransferase enzymes is often altered in cancer cells, resulting in the generation of glycans which are incompletely or aberrantly elaborated, or prematurely sialyated, such as the T-, T_N- and sialyl-T_N-antigens. Altered glycan expression can also be associated with autoimmune diseases.[23,25]

The density of α-linked GalNAc residues in mucin domains restricts the conformational freedom of the polypeptide backbone and results in the adoption of extended rod-like

glycoprotein structures which can be hundreds of nanometers in length.[8,23] Their dense clustering enables the glycans of mucin domains to shield the polypeptide backbone, thus enhancing the stability and serum half life of glycoproteins, and allows the glycoprotein to exhibit lubricant properties due to the hydration capacity of the oligosaccharides, for example in the intestinal lubricant glycoprotein MUC2.[23] Mucin glycans have also been shown to be important for the adoption of the correct protein conformation for signalling properties and binding. For example, mucin domains are necessary for the T-cell surface glycoprotein CD45 to adopt the correct monomeric conformation for binding to the protein kinase Lck and the initiation of T-cell activation.[23]

Mucin type glycans also function as multivalent ligands for the cell surface receptors which initiate cell-cell adhesion interactions. The most well-studied interactions of this type occur in the inflammatory response. The adhesion and "rolling" of leukocytes at sites of inflammation is controlled by the interaction of the selectin family of lectins (L(leukocyte)-, P(platelet)- and E(endothelial)-selectin) with sulfated glycans displayed by mucin glycoproteins on the surface of endothelial and leukocyte cells, such as PSGL-1 (P-selectin glycoprotein ligand-1).[8,26]

O-Linked glycans also function as ligands in a variety of other cellular adhesion events including those involved in human reproduction and viral and microbial infection.[8,23] Mucin type glycosylation is also found in the "antifreeze glycoproteins", mucin-like polymers composed of repeating *O*-glycosylated tripeptide units, which are found in

some families of Antarctic fish. The *O*-glycans displayed by these glycoproteins inhibit the growth of ice crystals and enable the fish to survive sub-zero temperatures.[27]

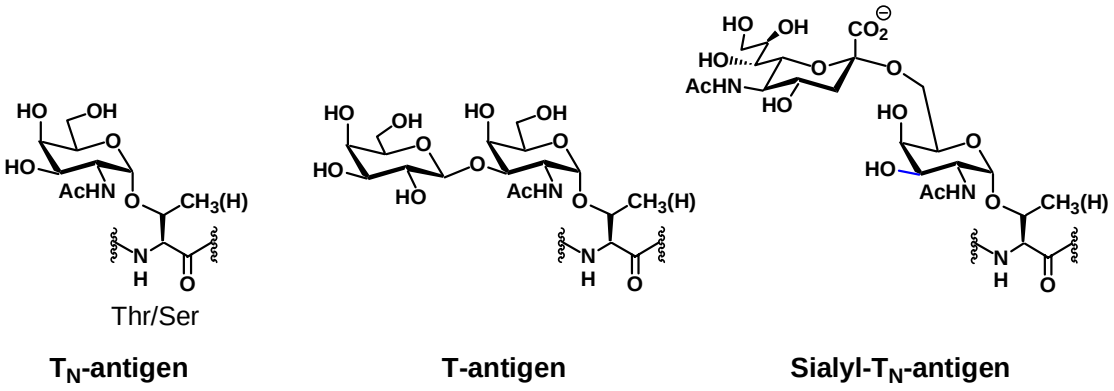


Figure 6: tumour associated antigens.

Glycoproteins expressed on the surface of cancer cells display epitopes known as "tumour associated antigens" (see **Figure 6**). Many of these antigens are truncated forms of mucin type *O*-glycans, such as the T_N- T- and Sialyl-T_N antigens, and can be specific for different types of cancer. For example, the T_N-antigen is typically found in human epithelial tumours, whereas the the sialyl-T_N-antigen is characteristic of breast, prostate and ovarian cancer. As these antigens can be recognised by the immune system, synthetic mucin type tumour associated glycopeptides are under investigation as anti-cancer vaccines.[25,28]

2. Assembly of glycoproteins and glycoprotein mimics

The defined, homogeneous glycoproteins required for biological studies cannot be practically produced in sufficient quantities by mammalian cells due the problem of

microheterogeneity, whereas recombinant proteins expressed by prokaryotes are not glycosylated. Yeast and fungi do have the ability to glycosylate proteins, but the glycans displayed are of non-human type and hence the resulting glycoproteins are immunogenic in humans. These problems have led to the development of several different strategies for the production of homogeneous glycoproteins. These include both biological and chemical synthetic techniques and strategies which involve elements of both approaches.

2.1 Glycoproteins from biological sources

Due to the aforementioned difficulties in the expression of glycoproteins, a large amount of research has been focussed on the development of advanced biological techniques for the generation of homogeneous glycoproteins. In "glycoprotein remodelling" approaches, the heterogeneous glycans expressed in mammalian glycoproteins are selectively trimmed down by enzymatic degradation. Glycosidase and endoglycosidase enzymes have been employed for this purpose, to cleave specific glycosidic bonds, or remove large oligosaccharide structures, respectively. Once the glycan has been reduced to a known core structure, often a monosaccharide or disaccharide unit, it can be used as the basis for the construction of a new, defined, oligosaccharide by enzymatic synthesis. This approach has been used to generate homogeneous glycoforms of proteins including ribonuclease B bearing sialyl Lewis X[29] and high-mannose type[30] glycans.

Much effort has also been focussed on the "re-engineering" of protein expression systems with the goal of producing defined homogeneous glycoproteins by altering their native glycosylation machinery. These techniques involve the incorporation of non-native glycosyltransferase enzymes into the expression system and the inhibition or removal of native enzymes through genetic manipulation.[31,32] The processes involved in glycan assembly are far from straightforward however, and have proved difficult to modify successfully. Despite these problems, promising advances have been made in this area. Gerngross and co-workers recently achieved the "humanization" of *N*-linked glycosylation in the yeast *Pichia pastoris* by employing large combinatorial genetic libraries and screening hundreds of constructs for activity at each stage of the "humanization" process. The resulting yeast could successfully produce proteins displaying largely homogeneous *N*-linked glycans and illustrates the potential for future development in the expression of homogeneous glycoproteins in fungal expression systems.[33]

A different approach to the incorporation of glycans during protein biosynthesis is based on the use of misacylated tRNAs which can site-specifically incorporate non-natural amino acids during translation in response to readthrough of a nonsense codon. ^{AUC}tRNA is synthetically acylated with the non-natural amino acid, which is then incorporated into the protein during expression in *E. coli* in response to the UAG codon. This method has been used to site-specifically incorporate glycosylated amino acids such as GlcNAc(α-*O*)Ser into proteins (see **Figure 7**). Other non-natural amino acids have also been incorporated into recombinant proteins using this technique.

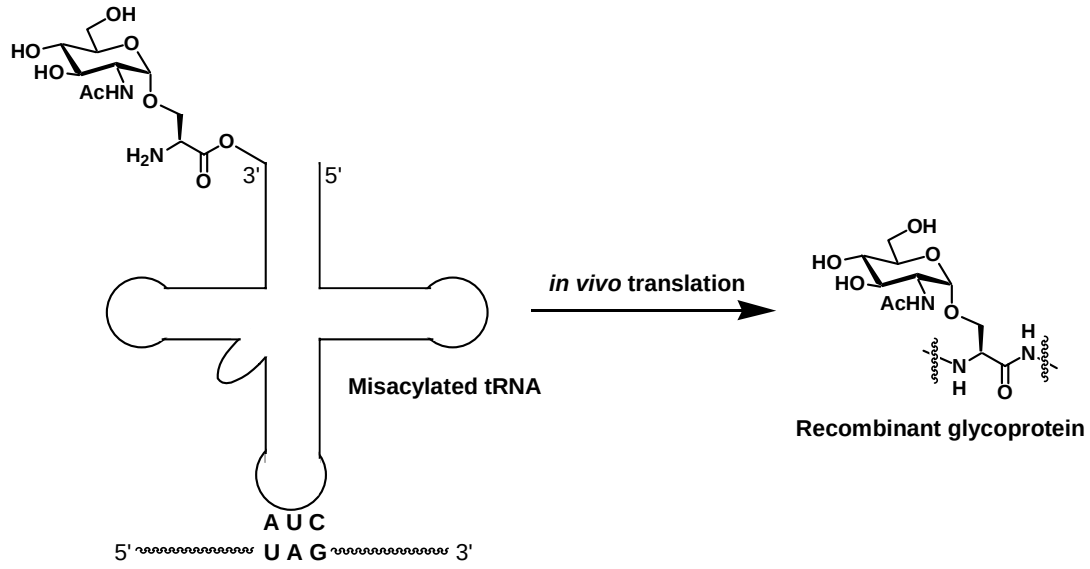


Figure 7: incorporation of GlcNAc(α-*O*)Ser into recombinant protein by misacylated tRNA

Recent work in the groups of Wong and Schultz has extended this methodology,[34,35] removing the need for the difficult synthesis of tRNA-amino acid conjugates. Their approach involves the evolution of an orthogonal nonsense supressor tRNA synthetase-tRNA pair to site-specifically insert an unnatural amino acid during protein biosynthesis. A library of tyrosyl tRNA synthetase active site mutants was evolved to generate a mutant tyrosyl tRNA synthetase(TyrRS)-tRNA pair which is capable of specifically charging the corresponding Amber-supressor tRNA ($mutRNA_{CUA}$) with the desired amino acid. The Amber stop codon TAG was introduced to the gene of a protein of interest and the unnatural amino acid was then site-specifically incorporated into the recombinant protein during translation in response to the Amber codon. tRNA synthetase-tRNA pairs have been evolved for the incorporation of GlcNAc(α-*O*)Ser and GalNAc(α-*O*)Thr in recombinant proteins expressed in *E. coli*. When a mutant

recombinant myoglobin protein, the gene of which was mutated to include the TAG codon, was coexpressed with the evolved TyrRS and mutRNA$_{CUA}$ genes in *E. coli* in the presence of the correct glycosylated amino acid, the glycosylated amino acid was expressed at the encoded site in the recombinant protein as monitored by mass spectrometry, although production levels were relatively low. Incorporation of GlcNAc(α-*O*)Ser was more efficient, as some background tyrosine expression was found for GalNAc(α-*O*)Thr incorporation.

Although these biological methods for glycoprotein generation represent impressive achievements and have undoubted potential in the future development of glycoprotein assembly, none have found general use due to the inherent drawbacks in their approaches. Many of the multiple specific enzymes necessary for glycoprotein remodelling are not commercially available or make use of expensive glycosyl nucleotide donors. The humanized yeast expression systems and the mutant tRNA synthetase-tRNA pairs are also not yet available, and the largescale biological effort involved in the evolution of such systems is beyond the scope of most chemistry laboratories. Chemical synthesis, either in total synthesis approaches or in partnership with enzymatic and semi-synthetic techniques, remains the only general method to access defined homogeneous glycoproteins.

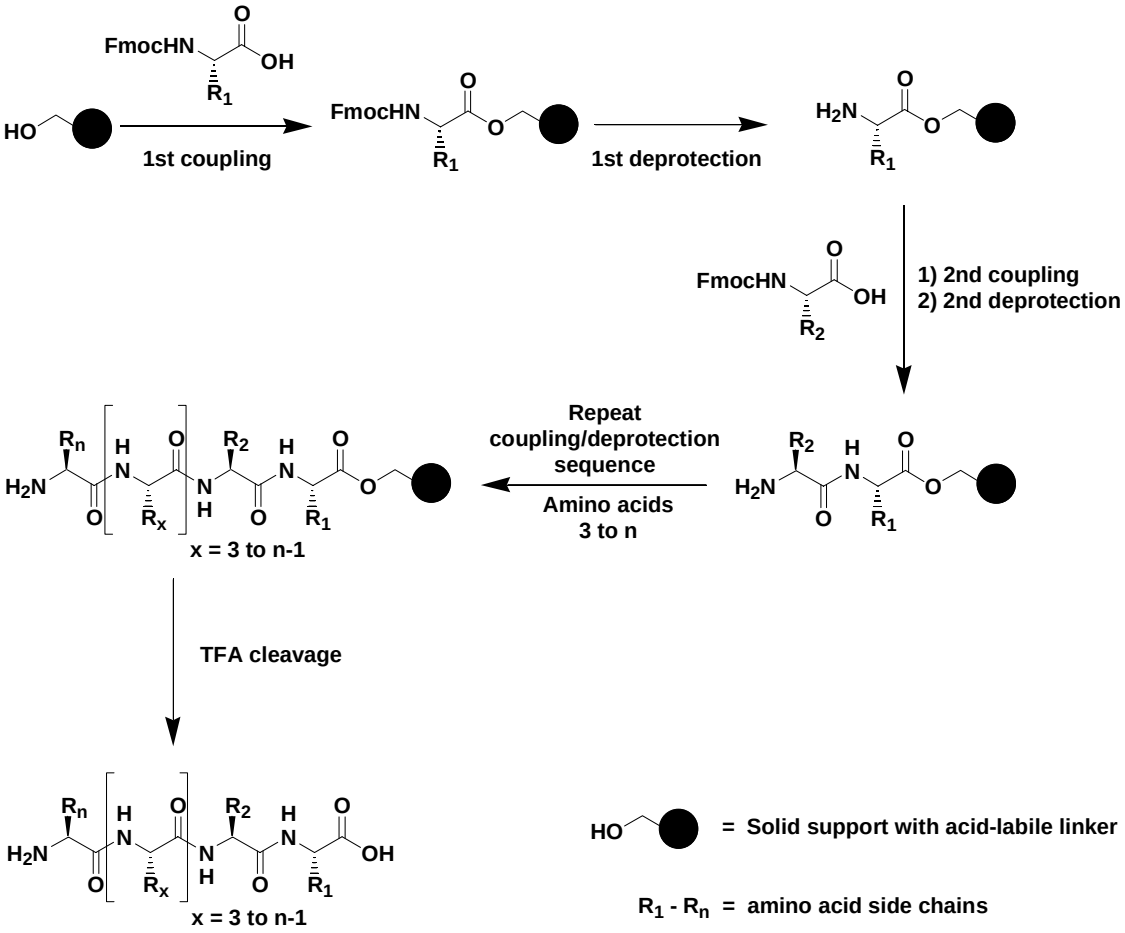
2.2 Synthesis of glycopeptides and glycopeptide mimics

The challenge of glycopeptide synthesis is typically approached using one of two different general strategies. Appropriately protected glycosylated amino acid "cassettes" can be synthesised and used in standard stepwise peptide synthesis. Alternatively, complex oligosaccharides can be coupled to synthetic peptides through natural or unnatural linkages in a convergent approach. The oligosaccharides can be introduced as fully elaborated glycans, or as simpler structures which can then be used as the basis for chemical or enzymatic elaboration following synthesis of the glycopeptide. Both strategies involve three main parts: peptide synthesis, oligosaccharide synthesis, and the formation of the key glycosidic linkage between the saccharide and either a single protected amino acid in the cassette strategy, or a polypeptide or protein in the convergent strategy.

2.2.1 Peptide Synthesis

Solid phase peptide synthesis (SPPS) is the standard technique for the synthesis of polypeptides.[36-39] Peptides are built up one residue at a time from the *C*-terminus by suitably protected amino acid building blocks, in a stepwise coupling-deprotection-coupling process (see **Scheme 1**). Following synthesis, the peptide is cleaved from the solid support and amino acid side chain protecting groups are removed to generate the native peptide. The vast spectrum of materials, reagents and methods available in SPPS, including the large number of solid supports, the wide range of differently labile

linkers and protecting groups and numerous coupling protocols makes it an extremely flexible and efficient synthetic tool. The use of automated SPPS is now also widespread and gives rapid access to large synthetic peptides.



Scheme 1: Fmoc solid phase peptide synthesis[39]

Two different protecting group strategies are commonly employed in SPPS. The Fmoc approach utilises amino acid building blocks protected by the *N*-9-fluorenylmethoxycarbonyl amine protecting group, which can be removed by nucleophilic cleavage with piperidine following coupling, and normally involves acid

labile linkers and side chain protection (see **Scheme 1**). The alternative older approach uses the *t*-butoxycarbonyl (Boc) amine protecting group, which is removed by acid treatment with trifluoroacetic acid (TFA) and employs linkers and side chain protection which must be subjected to hydrogen fluoride treatment to effect cleavage.

2.2.2 Oligosaccharide Synthesis

The oligosaccharide components of glycopeptides are considerably more challenging to synthesise than the peptide components due to their complex, branched structures, multiple, similarly reactive hydroxyl groups and the variable stereochemistry of the glycosidic linkage. The synthesis of oligosaccharides therefore requires both regiochemical and stereochemical control of the key glycosidation coupling reactions. Regiochemical control is normally achieved by the use of orthogonal protecting group strategies to allow the selective unmasking of the desired hydroxyl group.

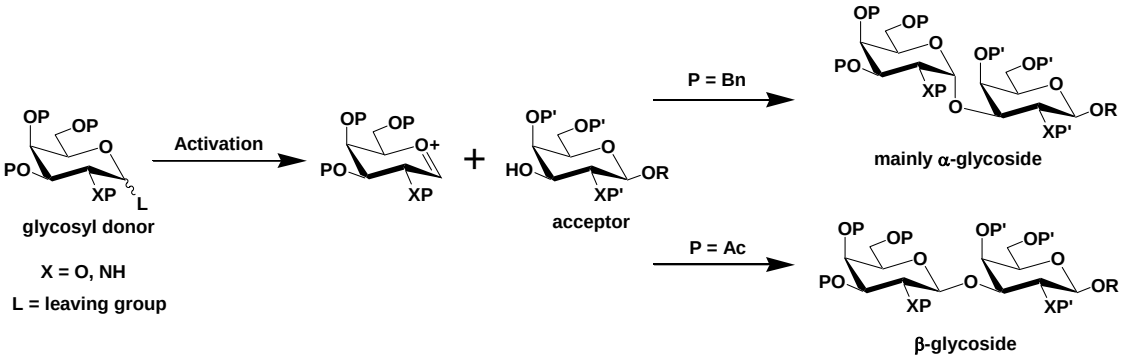


Figure 8: general chemical glycosidation reaction scheme.[5]

Many different methods exist for the formation of the glycosidic linkage, but typically a glycosyl donor saccharide with a leaving group at the anomeric position is activated, normally in the presence of a Lewis acid or heavy metal salt, to form the oxonium cation, which then reacts with the selectively exposed hydroxyl group of the glycosyl acceptor (see **Figure 8**). The stereochemistry of the glycosidic linkage can likewise be influenced in several ways, the simplest of which is to exploit the inherent preference for the formation of the α-glycoside due to the anomeric effect. Alternatively, the use of a protecting group on the neighbouring C2 substituent of the donor capable of neighbouring group participation or "anchimeric assistance" results in formation of the 1,2-*trans* glycoside (e.g the β-glycoside in **Figure 8**). Other methods of influencing anomeric stereochemistry include "intramolecular aglycon delivery" (IAD)[40] in which the donor and acceptor are tethered together in the correct orientation, or controlling the steric surroundings of the anomeric centre to favour the desired conformation.[41] Specific combinations of reagents and reaction conditions which favour a particular anomer can also be used.[42-44]

Despite the great advances in carbohydrate chemistry since the report of the first practical glycosidation method, the Koenigs-Knorr glycosidation, in 1901,[45] glycosidation reactions are far from trivial and require careful selection of the protecting groups on both donor and acceptor as well as the choice of leaving group, reagents and reaction conditions for the desired results to be achieved.[42,43] The synthesis of oligosaccharides is hence extremely labour intensive and time consuming due to the many protecting group manipulations and resulting purification steps necessary for the

formation of correctly reactive monosaccharide building blocks and their assembly into complex oligosaccharide structures.

Oligosaccharide synthesis represents a vast, complex and rapidly evolving area of research and therefore a detailed discussion of this field is beyond the scope of this report. Such details can be found in recent reviews.[42-44,46,47] In the context of glycopeptide synthesis however, it should be noted that oligosaccharides relevant to glycoproteins have now become synthetically accessible, through the efforts of several different research groups.[8,46] In some cases, innovations in carbohydrate chemistry such as solid phase and automated carbohydrate synthesis,[42,48,49] or reactivity-based one-pot methodologies[43,50-52] have aided these endeavours.

The use of chemo-enzymatic techniques is also becoming increasingly widespread in the synthesis of oligosaccharides.[52] The regio- and stereospecificity of enzymatic glycosidations, which remove the need for complex protecting group manipulations, and the mild reaction conditions under which they occur are obvious advantages over traditional chemical glycosidation reactions. Two major families of enzymes have been used to effect glycosidations: glycosyltransferases and glycosidases. Glycosyltransferases catalyse the transfer of a monosaccharide from a sugar nucleotide donor to saccharide acceptor with excellent stereo- and regiospecificity (see **Figure 9**), whereas the hydrolytic glycosidase class of enzymes can transfer monosaccharide or oligosaccaride donors to saccharide acceptors under either thermodynamic or kinetic control.

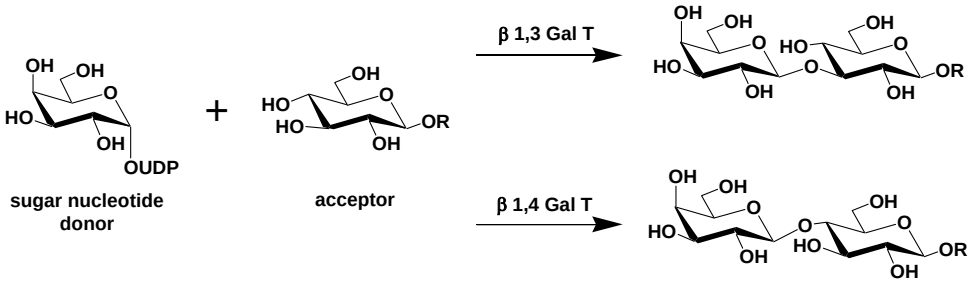
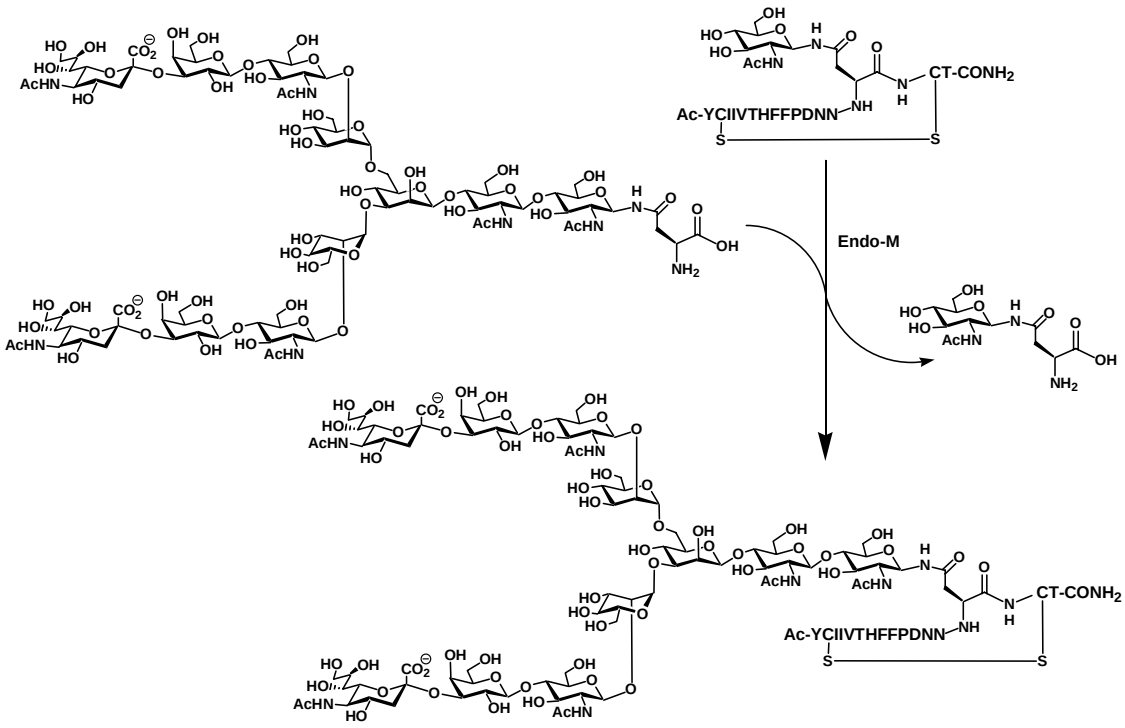


Figure 9: glycosyltransferase catalysed glycosidation reactions. The regio- and stereoisomers formed are dependent on the specific enzyme used.[5]

Another class of enzymes which is increasingly being employed in the synthesis of *N*-linked glycopeptides is the endoglycosidases. In nature, these enzymes recognise specific *N*-linked glycan chains and cleave the glycosidic linkage between the two GlcNAc residues in the pentasaccharide core, leaving only a single GlcNAc residue conjugated to the peptide backbone. The use of organic solvents and a large excess of glycosyl acceptor (the GlcNAc glycopeptide) causes the enzyme to catalyse a transglycosidation reaction which can transfer a large oligosaccharide corresponding to the normal glycan substrate to the GlcNAc glycopeptide acceptor (see **Scheme 2**).[8,53]

Recent work in this area has extended the scope of the transglycosidation reaction with the discovery that sugar oxazolines are effective donor substrates for the endoglycosidases. This improves the efficiency of the reaction and removes the need to incorporate the extra GlcNAc residue, wasted as a byproduct in previous procedures, into the donor glycan.[53-55]



Scheme 2: synthesis of glycosylated fragment of the nicotinic acetylcholine receptor via transglycosidation reaction catalysed by endo-β-*N*-acetylglucosaminidase Endo-M.[7]

In common with chemical approaches, the use of solid phase enzymatic techniques has also been reported in oligosaccharide synthesis, where either the saccharides or enzymes can be attached to the solid phase.[43] Also noteworthy is the recent work by Kajihara and co-workers in the semisynthesis of oligosaccharides.[56] This group has developed a method for the production of defined complex type *N*-linked oligosaccharides from egg-yolk. The oligosaccharides are obtained as short glycopeptides. Enzymatic treatment with a peptide:*N*-glycosidase enzyme cleaves the amide linkage between the peptide backbone and the oligosaccharide, which can then be used in the synthesis of glycopeptides and glycopeptide mimcs.

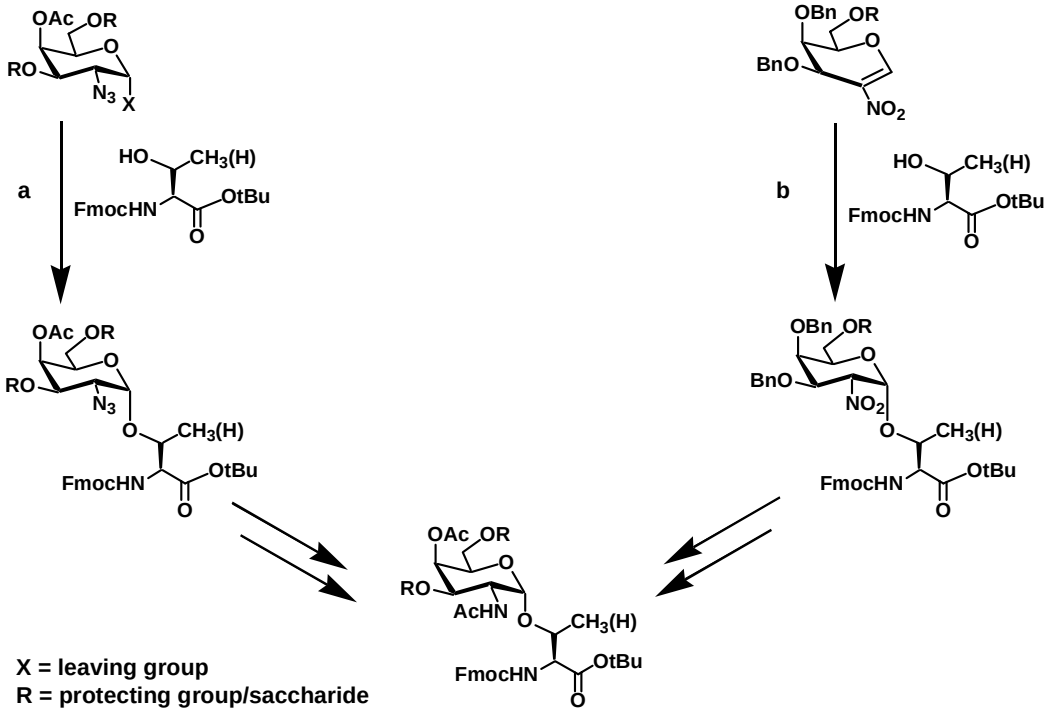
The disadvantages of enzymatic techniques in oligosaccharide synthesis are mainly associated with the availability of the enzymes and substrates. Enzymes for all the transformations which carbohydrate chemists might wish to perform are not available, and many of those which are are expensive and/or require the use of expensive reagents. Those enzymes which have a high tolerance for different substrates, such as the glycosidases, also suffer from a corresponding decline in efficiency and regioselectivity. In addition to this, some commercially available enzymes are only partially purified and can be contaminated with proteases which damage peptide chains. Nevertheless, enzymatic techniques are powerful and useful tools in the assembly of oligosaccharides and glycoproteins whose use is likely to grow still further in the future, especially when used in combination with chemical synthesis in chemoenzymatic approaches.

2.2.3 Formation of the oligosaccharide-peptide linkage

A variety of methods have been used to form the *O*- and *N*-glycosidic linkages between the peptide and oligosaccharide portions of glycopeptides and a still more diverse range of approaches is available for the formation of unnatural saccharide-peptide linkages.

The attachment of *O*-linked saccharides to peptides is normally accomplished by a typical glycosidation reaction, (see **Figure 8**, page 21) in which the hydroxyl side chain of serine or threonine acts as the glycosyl acceptor. For mucin type *O*-glycosidation, the most common type in nature and correspondingly the most commonly formed in synthesis, serine or threonine is glycosylated with *N*-acetylgalactosamine, or a larger

saccharide based upon this basic T_N antigen core. This is normally achieved by the glycosidation of the suitably protected amino acid with a 2-azido-galactosyl saccharide activated by a Lewis acid, or a soft electrophile or heavy metal salt (see **Scheme 3**). Traditionally, galactosyl bromides have been favoured, although a variety of other common leaving groups have also been used.[23,25] The 2-azido substituent, introduced to favour the formation of the desired α-anomer, can then be reduced to the amine and acetylated to give the desired GalNAc structure.



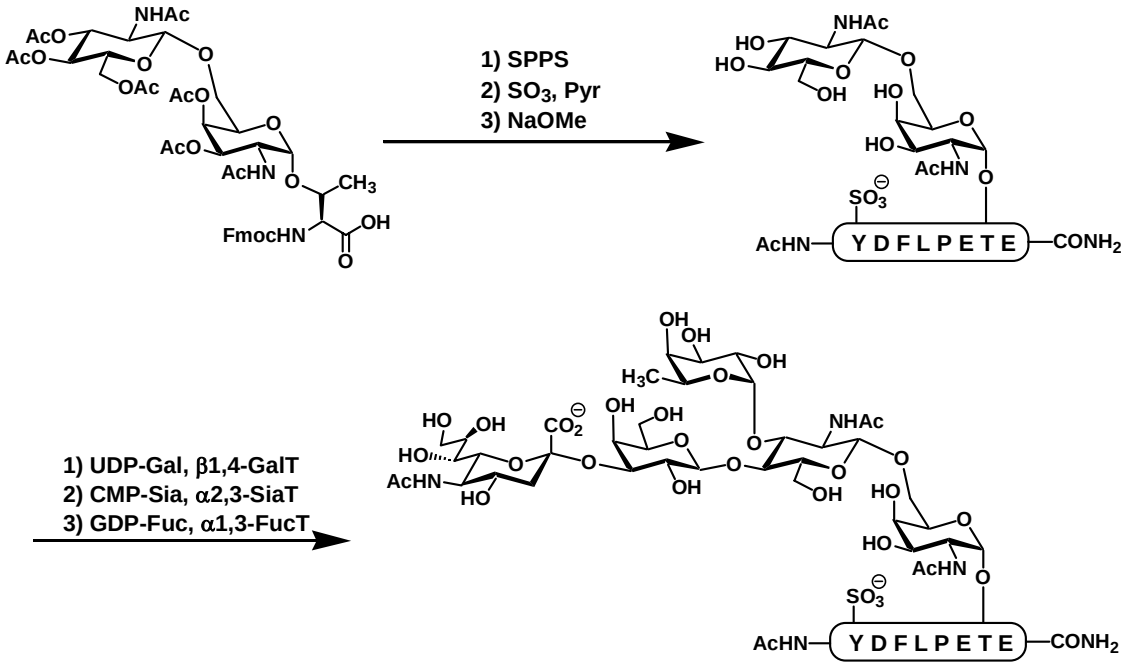
Scheme 3:[23] synthesis of glycosylated amino acid building blocks for assembly of mucin-type *O*-linked glycopeptides. a) glycosidation via leaving group activation; b) "nitroglycal concatenation" approach.

An alternative approach is the "nitroglycal concatenation" method developed by Schmidt and co-workers.[57] This method is based on the Michael addition of the hydroxyl side-chain of the relevant protected amino acid to suitably protected 2-nitroglycal to give the α-glycoside with high stereoselectivity. Reduction of the 2-nitro group and protecting group manipulation yield the desired GalNAc-Ser/Thr product (see **Scheme 3**).

In both cases the amino acid can be glycosylated with monosaccharides or with larger saccharide units. The glycosylated amino acid can then be further elaborated either chemically or enzymatically to produce complex glycan structures. The completed glycosylated amino "cassette" can then be used in standard solid phase peptide synthesis.[8] It should be noted however that additional complications arise from the use of saccharides in SPPS. The *O*-glycosidic linkages between monosaccharide units, and between saccharide and peptide, are acid labile and hence are not compatible with the repeated subjection to acidic conditions necessary for Boc-SPPS.[58] With hydroxyl groups correctly protected with electron-withdrawing protecting groups, typically acetyl or benzoyl esters, however, saccharides are compatible with the short TFA treatment used in Fmoc-SPPS to remove amino acid side chain protection. The use of strong bases is also incompatible with glycopeptide synthesis, as strongly basic conditions promote the β-elimination of the saccharide group from the serine or threonine side chain. The repeated piperidine treatments used to remove Fmoc protection are compatible with saccharides however, as piperidine is not usually a strong enough base to cause β-elimination.[53,59]

The strategy used for the assembly of the glycopeptide must also be chosen with care. Amino acids glycosylated with large oligosaccharides are often problematic to use in peptide coupling reactions, due to the large steric bulk of the glycan unit. They are also challenging and time consuming synthetic targets.[43] Chemical elaboration of the oligosaccharide following completion of the peptide chain is also very difficult due to the problems of monitoring reactions on solid phase and the small scales typically involved. The conceptually most efficient assembly strategy, a convergent coupling of the full size oligosaccharide with the completed peptide, has not been achieved for *O*-linked glycopeptides.

The most elegant and successful strategies used to date in the synthesis of mucin type glycopeptides use chemoenzymatic routes, involving the synthesis of an amino acid building block glycosylated with a small- to medium-sized glycan for use in SPPS. Following completion of the peptide chain, the oligosaccharide is elaborated enzymatically to form the desired glycan structure.[53] This approach is illustrated by the chemoenzymatic synthesis of a fragment of P-selectin glycoprotein ligand 1 (PSGL-1) by Wong and co-workers (see **Scheme 4**).[60] A threonine disaccharide building block was used in SPPS to create a monosulfated glycopeptide which was elaborated using commercially available enzymes to give the glycopeptide containing the desired pentasaccharide structure. A similar approach has recently been used by Nishimura and co-workers for the rapid synthesis of a larger MUC-1 glycopeptide bearing five *O*-linked hexasaccharide units in which several of the glycans are conjugated to adjacent amino acid residues.[61]



Scheme 4: chemoenzymatic synthesis of sulfate PSGL-1 fragment.[60]

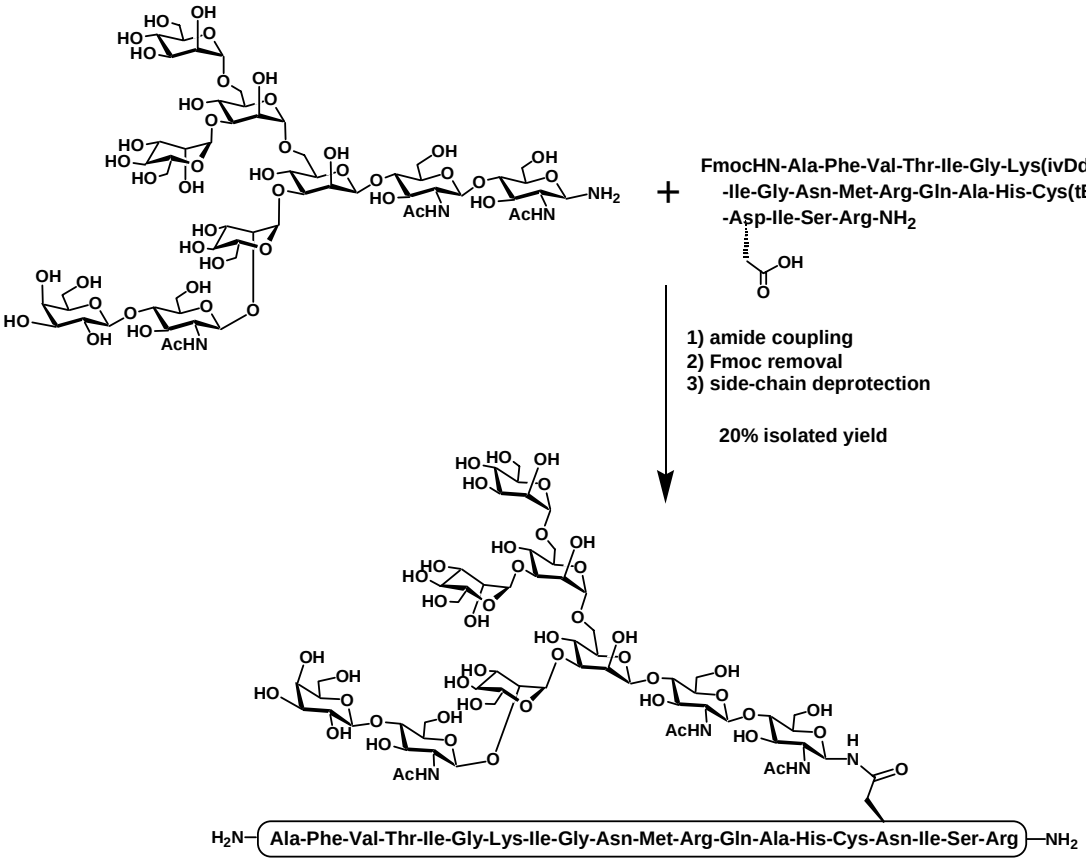
N-linked glycans are attached to the polypeptide backbone via an amide linkage to the asparagine side-chain, and therefore the methods for forming this linkage are different from those used for *O*-linked glycopeptides. The most straightforward and widely-used approach utilises a standard peptide coupling reaction between the carboxylate side chain of aspartic acid and the anomeric aminoglycoside of the appropriate glycan, which can be anything from a simple GlcNAc monosaccharide to large oligosaccharides based upon the common pentasaccharide core. The carboxylate coupling partner can be either the side chain of the apropriately protected aspartic acid, giving as the product the glycosylated amino acid "building block" for SPPS,[55,62-65] or a selectively exposed aspartate side chain in a preassembled peptide, in a convergent approach.[25,43,66-71] The aspartatic acid side chain has also been activated by conjugation to a 5-bromo-7-

nitroindoline group, which rearranges upon treatment with UV light to form an active ester.[72]

This peptide coupling reaction can be problematic in two particular respects, however. Firstly, the anomerisation of the aminoglycoside has been reported under peptide coupling conditions, to give a mixture of products containing the α-isomer as well as the desired β-glycoside.[73] Secondly, when a saccharide is coupled to an already-formed peptide chain, aspartimide formation can take place by intramolecular attack of the neighbouring amino group in the peptide chain on the activated carboxylate group.[6,74] It appears that anomerisation can be suppressed by the correct choice of coupling conditions, however.[14,74] For example, Danishefsky and co-workers have reported excellent results using the "Lansbury aspartylation"[69] of fully deprotected glycans.[70,74] The use of an excess of the peptide, whilst not preventing aspartimide formation, ensures the presence of sufficient peptide to give the maximum yield of coupled product, based upon the typically more valuable saccharide.[74] Interestingly, aspartamide formation has also been observed in the "building block" approach, via intramolecular attack of the amide linkage nitrogen on the activated ester during coupling of the building block in SPPS. Carefully optimised conditions for this coupling were subsequently developed to supress the side reaction.[65]

The applicability of both the "building block" and the convergent assembly approaches allows greater flexibility in the synthesis of *N*-linked glycopeptides than is the case for their *O*-linked counterparts. Both strategies have been successfully employed, but the more elegant convergent approach is the most common strategy for the chemical

synthesis of *N*-linked glycopeptides, due to the previously mentioned inhibition of peptide coupling yields by large glycan structures.[43,64,71] The sensitive glycan is also exposed to fewer reactions in this approach and in fact saccharides can be used without hyroxyl group protection. The method is particularly useful for large glycan structures as it uses less of the valuable oligosaccharide than the building block approach. The power of the convergent approach has been illustrated by several impressive total syntheses of peptides glycosylated with large high-mannose and complex type *N*-linked oligosaccharides, for example fragments of the HIV surface envelope protein gp120, by Danishefsky and co-workers (see **Scheme 5**).[68,75]



Scheme 5: total synthesis of glycosylated fragment of gp120.[68]

The problems detailed above, which can be associated with direct aspartylation of aminoglycosides, have led to the development of several alternative methods of forming the amide linkage between saccharide and peptide. Anomeric isothiocyanates of appropriately protected GlcNAc and larger saccharides can react with the carboxylate side chain of aspartic acid under mild heating conditions to form the desired amide linkage.[76] The Ritter reaction has also been used. Anomeric acetonitrilium intermediates, formed from the activation of pentenyl glycosides in the presence of acetonitrile, will react with the carboxylate side chain to give the amide linkage following rearrangement.[77] A 2-phthalimido substituent is necessary in this case to ensure the β-nitrilium intermediate is formed by neighbouring group participation. The glycosidation of silylated amide side chains of asparagine has been achieved with glycosyl sulfoxide donors.[78] These methods have seen relatively little use in recent years, however.

More recent work has investigated the use of the Staudinger ligation between glycosyl azides and phosphinothioester derivatives of aspartic acid to form the β-amide linkage with high stereoselectivity. Yields were moderate, however, for the ligation of even relatively simple glycosyl azides.[79] Davis and co-workers have also recently reported the use of a three component Staudinger ligation between the glycosyl aza-ylide intermediate, formed by the reaction of a glycosyl azide with tributylphosphine, and an activated aspartate side chain, to form the β-amide linkage.[80] Unprotected mono- and disaccharides were conjugated to suitably protected asparatic acids and a short model peptide in good yield by this technique.

Benzyl hydroxamate derivatives of asparagine have also been used as glycosyl acceptors in glycosidation reactions with glycosyl fluoride donors, to give the amide-linked saccharide in good yield and stereoselectivity following reductive removal of the benzyloxy group.[81] The use of this procedure for more complex systems than monosaccharides and single protected amino acid derivatives has yet to be reported, however. Direct glycosidation of the aspartamide side chain with glycosyl β-*N*-phenyltrifluoroacetimidate donors has also been reported in good yield and stereoselectivity for glycosylation of suitably protected asparagines and asparagine dipeptides with simple model monosaccharides.[82] In common with their *O*-linked counterparts, *N*-linked glycopeptides have also been assembled by chemoenzymatic routes using either glycosyltransferase[83] or endoglycosidase[7] enzymes for elaboration of the glycan structure following peptide synthesis.[8,71]

A wide range of unnatural peptide-saccharide linkages has also been used in the synthesis of glycopeptide mimics. These have been developed for a number of purposes, for example to form linkages more resistant to hydrolysis under physiological conditions, but are primarily used to achieve the chemoselective glycosylation of peptides.

Chemoselective glycosylation to form glycopeptide mimics has been used as an alternative to the synthesis of native glycopeptides, as these reactions typically take place under mild conditions in the presence of a variety of other functional groups, and are therefore particularly useful for sensitive and heavily functionalised molecules such as saccharides and peptides. In a typical chemoselective glycosylation, the saccharide is

modified with a functional group which uniquely recognises and reacts with a complementary group on the peptide. The converse approach, in which the peptide is functionalised to react selectively with the anomeric position of the saccharide, has also been used. In most cases these unnatural linkages appear to have little effect on the conformation and properties of their appended glycans.[84]

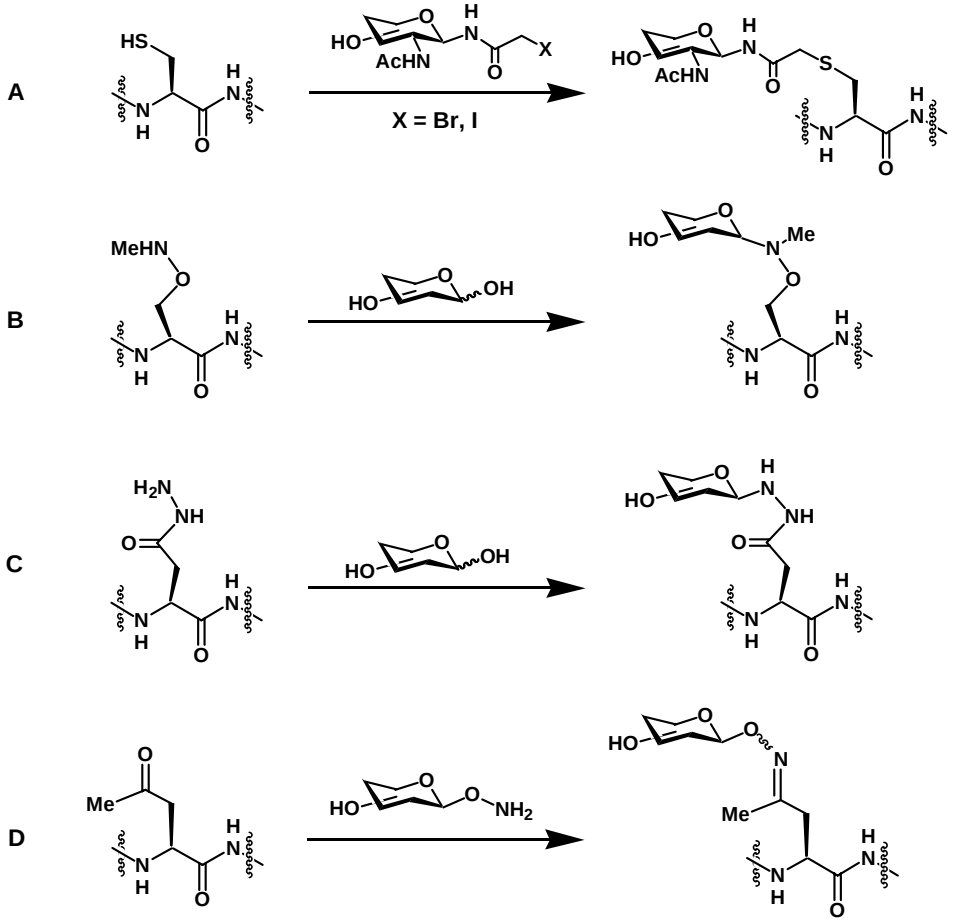


Figure 10: chemoselective glycosylations. Reaction of cysteine with glycosyl α-haloacetamide (**A**); reaction of hydroxylamine (**B**) or hydrazide (**C**) derivitised peptides with unprotected saccharides; and reaction of ketone derivitised peptide with glycosyl hydroxylamine (**C**).

The thiol side chain of cysteine has been used extensively as a handle for the chemoselective modification of proteins, as it is capable of reacting with saccharides functionalised with several different reaction partners such as α-haloacetamides[85] (see **Figure 10**) and bromoethyl saccharides[86] to form thioether-linked products, or thiosulfonates[87] and dithiopyridyl saccharides[88] to form disulfide-linked products.

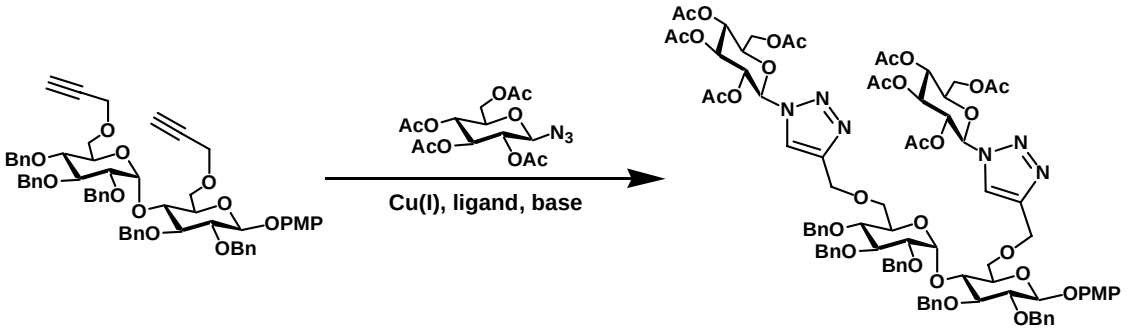
S-linked glycosyl amino acids and dipeptides for use in the synthesis of *S*-linked glycopeptide analogues have also been formed by direct glycosidation of cysteine derviatives with glycosyl halides,[89,90] or by the coupling of glycosyl thiolates with β-bromoalanine derivatives.[91]

The chemoselective reaction between a ketone and a hydroxylamine or hydrazine derivative has also been extensively used for the chemoselective glycosylation of peptides. Unnatural amino acids containing hydroxylamine[92, 93] or hydrazide[93] functionalised side chains can be incorporated into peptides by SPPS. These can then react chemoselectively with the reducing end of unprotected saccharides to give oxime- or hydrazide-linked glycopeptide mimics (see **Figure 10**).

The converse approach has again been used in the case of these reactions, in which the ketone is incorporated into the peptide by SPPS. The peptide can then be chemo- and site-selectively modified with functionalised saccharides such as glycosyl hydroxylamines (see **Figure 10**),[94,95] thiosemicarbazides[95] or hydrazides.[95]

Several other unnatural linkages have been used in the synthesis of glycopeptide mimics, although those detailed above represent the most commonly used examples.[6,84]

One chemoselective reaction which is becoming increasingly popular in the context of saccharide linkages is the Huisgen [3 + 2] cylcoaddition[96] between an azide and an alkyne to form a 1,2,3-triazole-linked species.[97,98] The development of this reaction by Sharpless and co-workers into a copper-catalysed "click chemistry" reaction has enabled it to be used at physiological temperatures and in heavily functionalised environments and indeed this approach has been used for the modification of a range of biomolecules (see **Scheme 6**).[99-105]



Scheme 6: example of "click chemistry" [3 + 2] cycloaddition reaction in formation of unnatural saccharide linkages.[97]

2.3 Assembly of Glycoproteins

The methods detailed in the previous sections, in particular the advances in oligosaccharide synthesis and chemoenzymatic methods for glycan synthesis and elaboration, have allowed access to increasingly large and complex glycopeptides and several impressive examples of the synthesis of naturally important glycopeptides, such as HIV-1 V3[55] and CD52,[15] or glycopeptide mimics have been reported, as previously described. A limitation in glycopeptide synthesis is the size of the polypeptide backbone, however. Only peptides up to approximately 50 amino acid residues in length can practically be produced by SPPS. Beyond this limit, low coupling yields and byproduct formation render further peptide elongation by this method impractical. The solution to this problem is peptide ligation: a chemoselective condensation between two peptide segments, each assembled by SPPS and either partially or fully deprotected. Such reactions typically take place under mild aqueous conditions and allow straightforward isolation of the larger ligation product from the smaller peptide coupling partners.

2.3.1 Native Chemical Ligation

The native chemical ligation (NCL) reaction, or native peptide ligation as it is also known, developed by Kent and co-workers is one of the most powerful and widely used methods for the assembly and modification of peptides and proteins.[106] The reaction takes place between two peptide coupling partners: one with an *N*-terminal cysteine

residue and the other with a *C*-terminal α-thioester. Nucleophilic attack of the cysteine side chain thiol leads to a reversible transthioesterification, which is followed by an rapid irreversible *S* to *N* acyl-shift via a five-membered ring transition state to form the native peptide linkage (see **Figure 11**). The reaction is entirely chemoselective and is carried out at ambient temperatures under aqueous conditions, and is therefore compatible with fully unprotected peptides and saccharides.

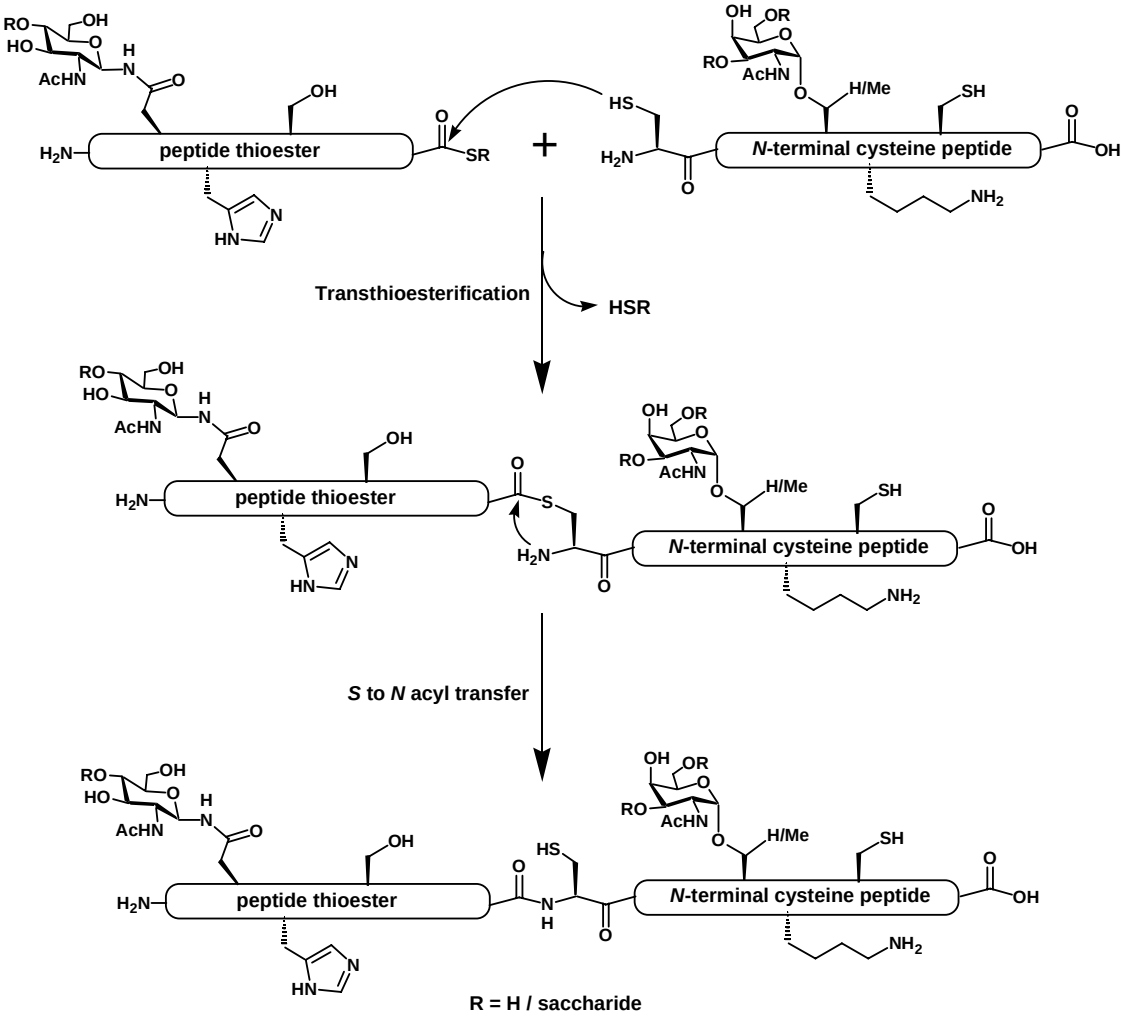


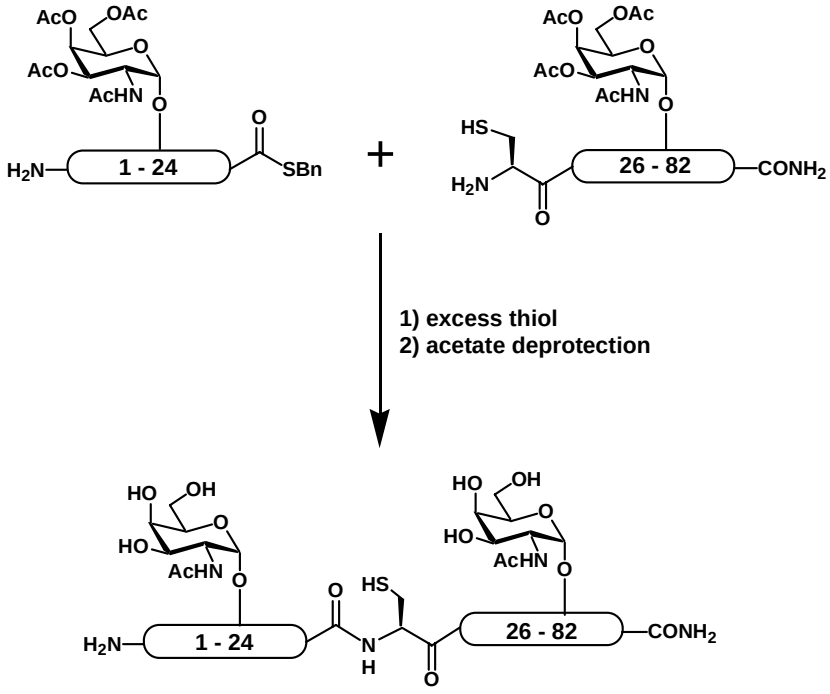
Figure 11: native chemical ligation of unprotected glycopeptide segments.

Solubilizing agents such as guanidine hydrochloride or urea do not interfere in the reaction and can be used to allow higher peptide concentrations. The reaction progresses most efficiently under slightly basic conditions (pH = 7.0 – 8.0) which favour the initial nucleophilic attack without being strong enough to hydrolyse the thioester to the corresponding acid.

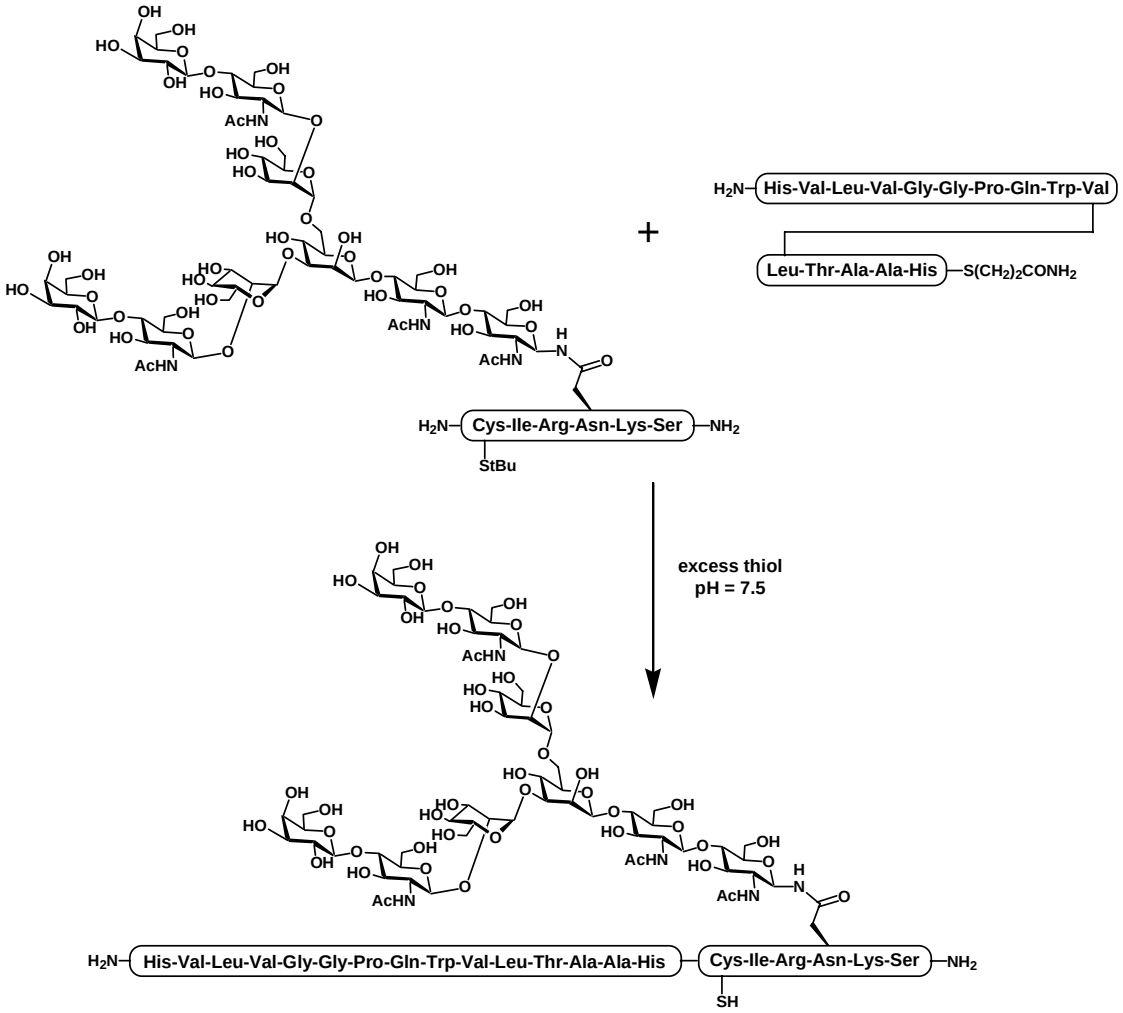
Addition of an excess of thiol, typically thiophenol or 2-mercaptoethylsulfonic acid sodium salt (MESNa) has also been found to increase the efficiency of the reaction by helping to prevent oxidation of the cysteine thiol. The excess thiol also participates in reversible transthioesterification with the thioester component. This has no negative impact on the reaction however, as the resulting thioester is also reactive towards the cysteine-peptide and the key intermediate thioester formed by the reaction of the two peptide coupling partners rearranges too rapidly to be broken down again by exchange with free thiol. In fact thioester exchange with the additional thiol can have a positive effect on the reaction by increasing the rate at which the unprofitable intermediate thioesters, which can be formed by the reaction of the thioester with internal cysteine residues in either peptide, or indeed the cysteine thiol formed at the ligation site following rearrangement, are broken down.[107]

Ligations are typically fast and efficient, although difficulties have sometimes been encountered with thioesters containing β-branched *C*-terminal amino acid residues such as isoleucine or valine.[106] The mechanism of NCL has recently been further investigated by Kent and co-workers.[107] In this study, the ability of a range of different thiol additives to catalyse the rate-determining transthioesterification step was examined

for a model ligation at a Leu-Cys junction. Aryl thiol additives with pKa > 6, which gave a good balance between reactivity for the thiol exchange reaction and the reactivity of the resulting thioester, gave the best results. The thiol exchange reaction was in fact found to be the rate-determining step for ligations in the presence of these additives, so rapid was the subsequent transthioesterification. Two thiol additives in particular, 3-hydroxythiophenol and 4-mercaptophenylacetic acid (MPAA), were very effective catalysts for ligation. In contrast, alkyl and benzyl thioesters, such as those formed from the commonly used ligation additives MESNa and benzyl mercaptan, respectively, were less reactive and underwent transthioesterification more slowly, although the initial thiol exchange reaction was more rapid. These interesting results should facilitate improvements in the efficiency of NCL.



Scheme 7: synthesis of diptericin by native chemical ligation.[108,109]



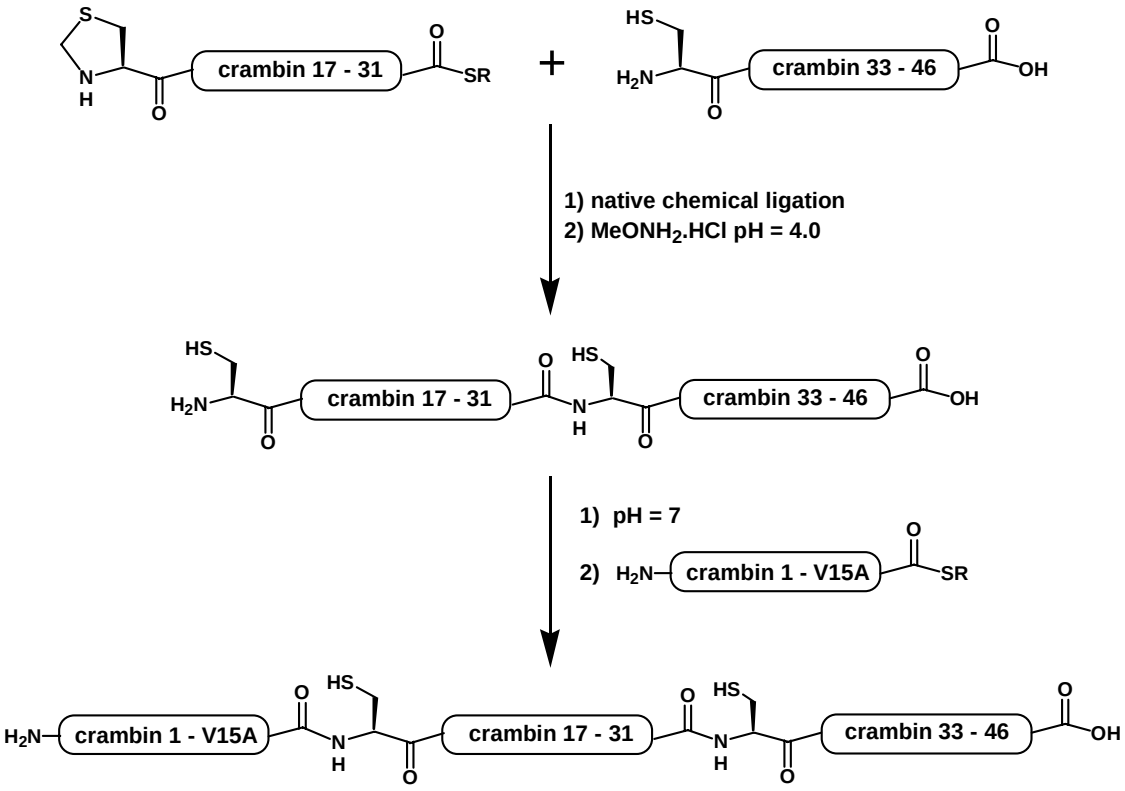
Scheme 8: synthesis of PSA glycoprotein by native chemical ligation.[67]

N-terminal cysteine peptides can be synthesised in a straightforward manner by SPPS. The synthesis of *C*-terminal thioester peptides however, has required the development of specialised approaches (see section **2.3.10**). NCL has been used in the assembly of numerous proteins and glycoproteins, such as the antimicrobial insect glycoprotein diptericin[108,109] and prostate specific antigen (PSA) glycoproteins[67] (see **Schemes 7** and **8**).

Native chemical ligation of two peptide components allows the assembly of proteins and glycoproteins of small to medium size. For larger targets, repeated ligations must be used in an iterative approach. This involves the use of one or more peptide fragments which contain both a *C*-terminal thioester and an *N*-terminal cysteine residue, the latter of which must be selectively masked with a protecting group removable under mild conditions compatible with the unprotected peptide. Following ligation, this protecting group is removed to allow the newly formed *N*-terminal cysteine peptide to participate in ligation with a further thioester peptide segment. The peptide must be isolated after each ligation and deprotection step, typically by semi-preparative high performance liquid chromatography (HPLC) and lyophilization and so the iterative approach can be time-consuming.

To overcome this problem, Kent and co-workers have developed a method for the "one-pot" total synthesis of proteins.[110,111] The approach utilises the 1,3-thiazolidine-4-carboxo (Thz) group to protect the *N*-terminal cysteine residue of the *N*-terminal cysteine/*C*-terminal thioester peptide segment (see **Scheme 9**). It is possible to remove this protecting group by treatment with methoxylamine hydrochloride at pH ≈ 4. Following complete ligation of an *N*-terminal cysteine peptide and an *N*-terminal Thz/*C*-terminal thioester peptide, the crude reaction mixture was accordingly treated with methoxylamine hydrochloride in this way to unmask the new *N*-terminal cysteine thiol group. The pH of the reaction mixture was then returned to approximately 7 by the careful addition of buffered sodium hydroxide and the second thioester peptide segment was added. The second ligation proceeded smoothly despite the presence of methoxylamine, which reacts only very slowly with thioesters at pH 7. This "one-pot"

approach has been successfully applied to the total chemical synthesis of the proteins crambin[111] and ubiquitin.[110] Danishefsky and co-workers have also recently reported the synthesis of glycopeptides bearing three oligosaccharides groups by an iterative ligation approach using the Thz protecting group.[112,113]



Scheme 9: "one-pot" total chemical synthesis of crambin.[111]

Kent and co-workers recently improved the speed and efficiency of the iterative assembly process still further by the incorporation of a histidine His_6 tag at the *C*-terminus of the *C*-terminal peptide segment.[114] This allows speedy and simple purification of the reaction mixture by Nickel affinity column following each ligation-deprotection step in an alternative approach to the "one-pot" method. The utility of this

approach was demonstrated by the synthesis of the 17 kDa modular repeat protein TPR, involving the iterative ligation of four separate peptide segments.

2.3.2 Expressed Protein Ligation

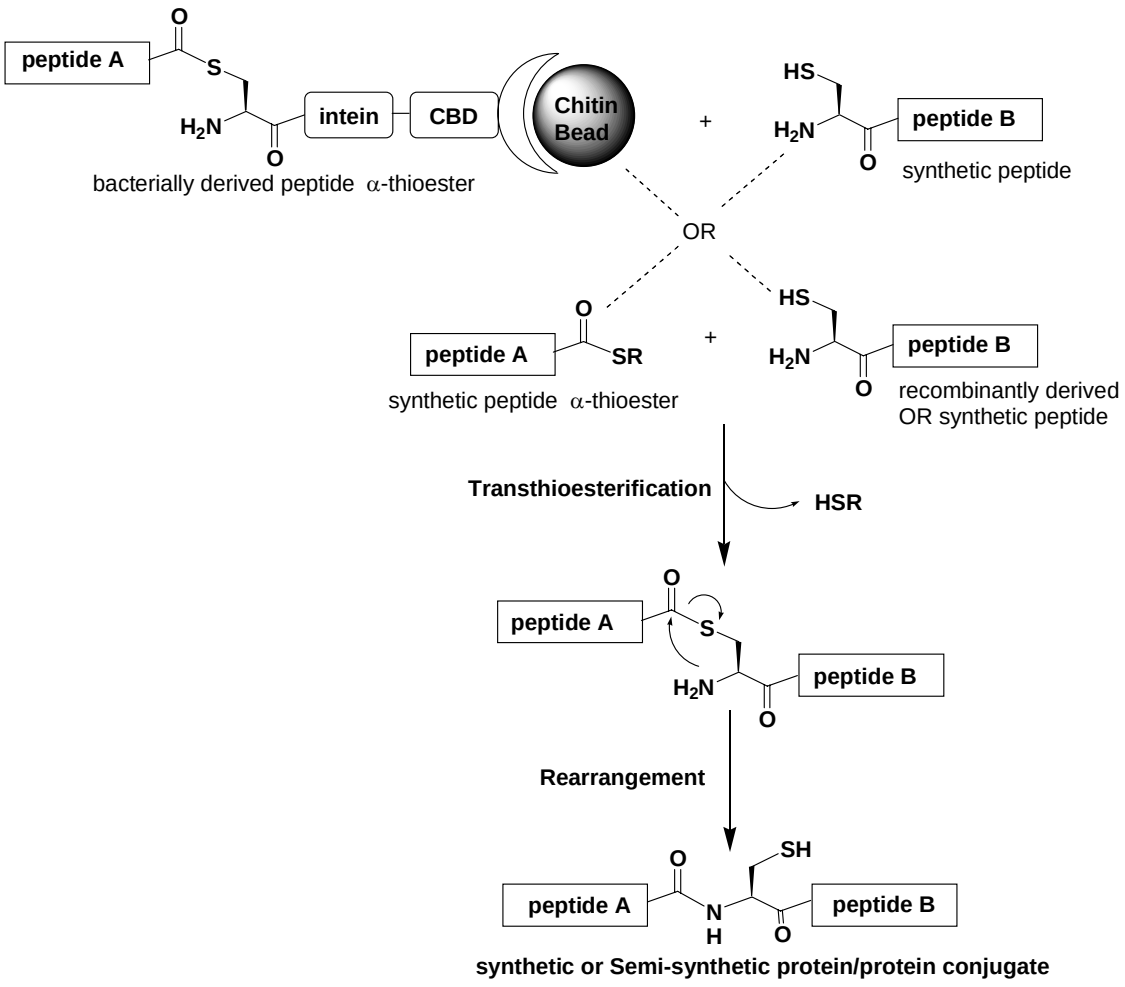


Figure 12: expressed protein ligation.

Despite these advances, most proteins and glycoproteins are still too large (≥30 kDa)[114] to be assembled by total chemical synthesis. These targets can be reached however, by

the use of semi-synthetic techniques. Protein semi-synthesis involves the fusion of synthetic and biologically expressed polypeptide fragments to form proteins. The primary method used for the coupling of synthetic and recombinant peptide fragments in protein semi-synthesis is expressed protein ligation (EPL).[115-117]

This technique, developed by Muir and co-workers, involves a native chemical ligation in which either the *N*-terminal cysteine segment or the *C*-terminal thioester segment is bacterially-derived (see **Figure 12**). As the size of recombinant proteins expressed by organisms such as *E. coli* is not limited by the constraints of SPPS, the assembly of large proteins containing synthetic segments is possible. Hence short synthetic glycopeptides can be fused to bacterially-derived protein fragments to form large, structurally homogeneous glycoproteins.[118] This approach has also been used to incorporate a wide variety of synthetic tags,[119] probes[120] and structural motifs[121] into biologically important proteins.

Recombinant *N*-terminal cysteine peptide fragments are typically expressed as fusion proteins in which the desired peptide sequence is conjugated to a purification tag via a specific sequence of amino acid residues which can be recognised and selectively cleaved by a commercially available enzyme such as factor Xa protease[118] or TEV proteases[122] to give *N*-terminal cysteine polypeptides. Recent work in our group has also shown the utility of cyanogen bromide for the cleavage of *N*-terminal cysteine polypeptides from recombinant fusion proteins.[123] This alternative approach is particularly valuable for fusion proteins which are insoluble under conditions compatible with the proteolytic enzymes mentioned above.

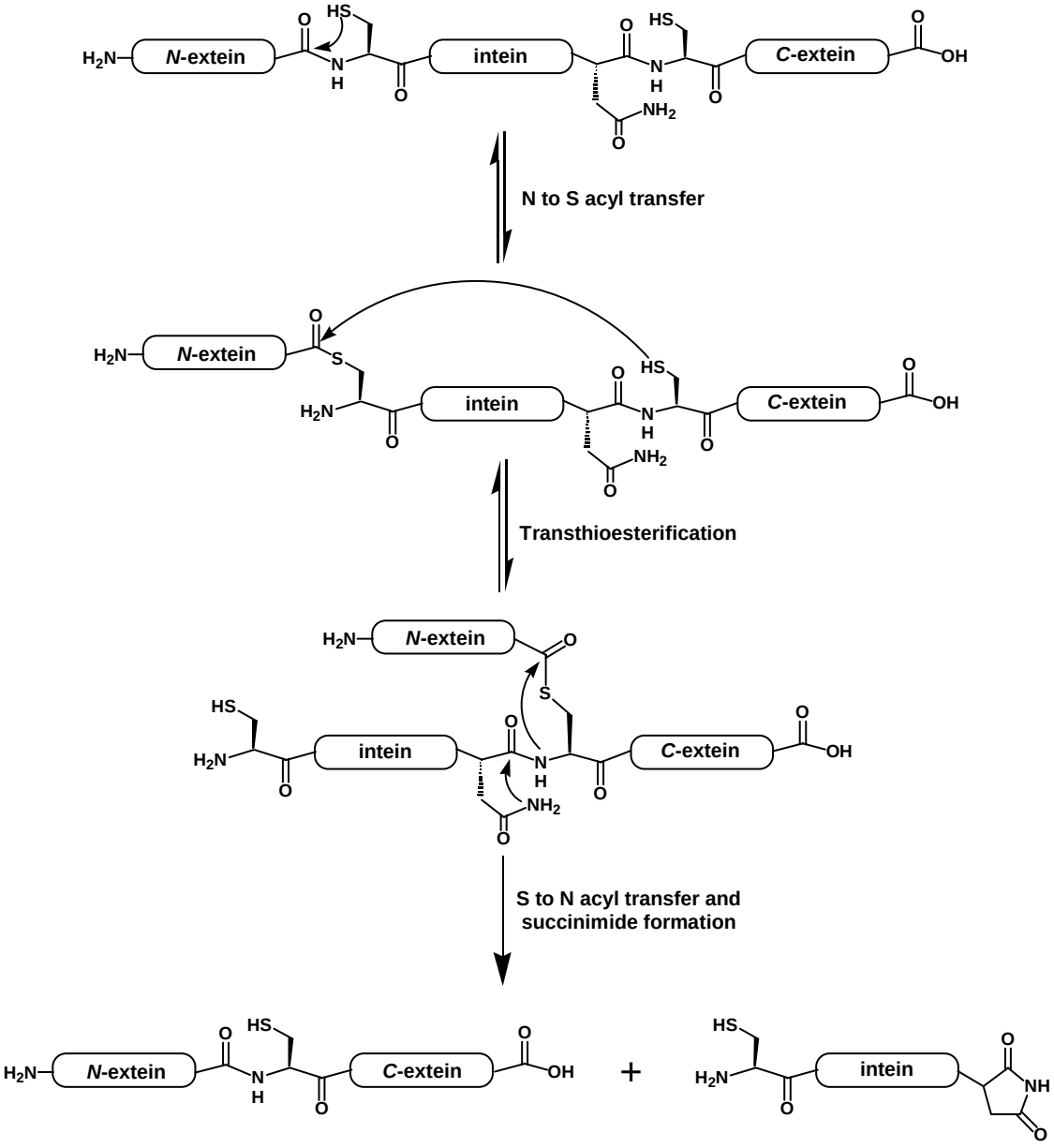


Figure 13: mechanism of intein-mediated protein splicing.

Recombinant polypeptide *C*-terminal thioesters can be produced as *C*-terminal fusion proteins with modified inteins, which rearrange to generate the thioester in an approach developed by Muir and co-workers.[115,116] Inteins are naturally occurring protein splicing elements which are excised from proteins via a series of intramolecular

rearrangements which result in the ligation of the flanking polypeptide sequences, or exteins (see **Figure 13**). It is possible that the intein domain itself initiates protein splicing by twisting the polypeptide backbone into a favourable conformation for the rearrangements to occur.[117] The fusion of the target polypeptide to a modified intein, which can promote only the first *N* to *S* acyl shift and cannot be excised, leads to the formation of a recombinant thioester which can then be released from the intein by intramolecular transthioesterification with a thiol or with an *N*-terminal cysteine peptide in native chemical ligation.

In recent work in our group the power and versatility of the NCL/EPL approach to glycoprotein assembly was demonstrated by the formation of structurally homogeneous glycoforms of the cellular adhesion glycoprotein GlyCAM-1.[118] Synthetic glycopeptides and glycopeptide thioesters were ligated to both *N*-terminal cysteine and *C*-terminal thioester polypeptides expressed in *E. coli* to produce three full-length, homogeneously glycosylated variants of the target protein.

2.3.3 Cysteine-free Peptide Ligation

The major limitation of the NCL/EPL approach is its reliance on the presence of the relatively uncommon cysteine residue at the ligation junction. Only about 1.7 % of all residues in known proteins are cysteines.[53] It is therefore quite probable that a given target protein or glycoprotein will either not contain a cysteine residue at all, or that a native cysteine will not be present at a site appropriate for use as a ligation junction.

Ligation junctions must obviously be sited at points where the ligation coupling partners will be available via chemical synthesis or protein expression, and, in the case of glycoproteins, cannot be directly adjacent to residues glycosylated with oligosaccharides, as the steric bulk of these groups prevents ligation.[68]

The introduction of non-native cysteine residues at strategic points to accommodate ligation can be problematic. Altering the amino acid sequence can affect the structure of the protein. The unnatural cysteine residue can also cause problems with any subsequent ligations. In the semisynthesis of GlyCAM-1 carried out in our group, for example, a non-native cysteine residue was introduced into one of the glycopeptide fragments to facilitate its assembly by NCL. It was found that the subsequent EPL of this fragment with a bacterially derived thioester proceeded extremely slowly, unless the non-native internal cysteine was capped by treatment with iodoacetamide to form a glutamine analogue. To extend the applicability and generality of the NCL/EPL approach and overcome problems such as this, several strategies for cysteine-free peptide ligation have been developed, using a range of different approaches.

2.3.4 Post Ligation Modification of Cysteine and Selenocysteine

A "desulfurization" approach can be applied to allow the formation of non-cysteine containing proteins by NCL.[124] In this technique NCL takes place as normal between a thioester and an *N*-terminal cysteine peptide. The ligation product is then subjected to hydrogenation in aqueous acid in the presence of aluminium oxide and palladium, or

treated with Raney nickel, to convert the cysteine residue into the much more commonly found alanine. The stereochemistry of the amino acid side chain is preserved in the reaction, which proceeded in generally excellent yield for the model peptides studied.

This procedure is not particularly general, as it is obviously incompatible with peptides or proteins containing native cysteine residues, which are also desulfurized. Lower yields for the reaction were found for peptides containing methionine due to peptide damage caused by slow desulfurization of these residues. Tyrosine phenol side chains can also be converted to cyclohexanone under the desulfurization conditions.[125]

N-terminal selenocysteine peptides have been shown to participate in NCL and EPL with thioesters in the same manner as their cysteine counterparts, although these ligations can give mixtures of products due to the tendency of the selenol side chain to react further with a second thioester molecule or with the thiol reagent present in the reaction mixture.[126,127] Selenocysteine can be reduced to alanine by catalytic hydrogenation under similar conditions to cysteine, although the same comments regarding damage to the peptide apply. Alternatively, mild oxidative treatment with hydrogen peroxide under aqueous conditions converts the selenol side chain to dehydroalanine, which can serve as a handle for further functionalisation by nucleophilic attack.[127]

2.3.5 Acyl Transfer Auxiliaries

An interesting and potentially more general approach to cysteine-free peptide ligation is the use of thiol acyl transfer auxiliaries which can be appended to the *N*-terminus of a peptide to mimic the action of the cysteine thiol side chain in NCL. Ideally, the auxiliary is cleaved following ligation, resulting in a "traceless" cysteine-free peptide ligation (see **Figure 14**).

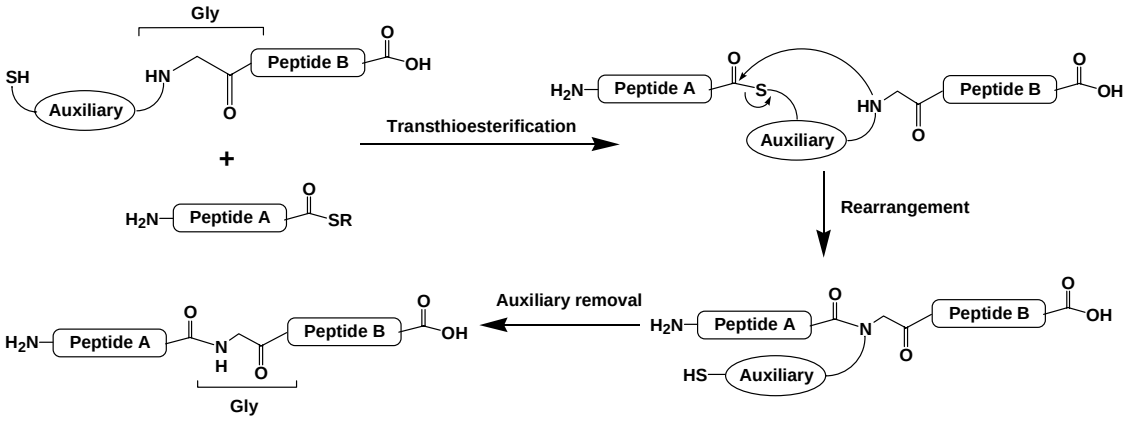


Figure 14: thiol auxiliary mediated cysteine-free peptide ligation at an Xxx-Gly junction.

Early examples of these auxiliaries were based on the simple modification of the *N*-terminal amino acid with an ethanethiol group to give **1** (see **Figure 15**).[128] This auxiliary functions in a very similar way to cysteine, with *S* to *N* acyl shift taking place by an intramolecular rearrangement via a five-membered ring transition state. Good results were obtained for ligation at a model Gly-Gly junction, although the auxiliary was not removable following ligation. The aminoxy-linked analogue **2**, designed to

allow reductive cleavage by treatment with zinc dust following ligation, showed a slower rate of reaction for the Gly-Gly junction and reacted only very slowly at the more sterically demanding Lys-Gly and Gly-Ala junctions.

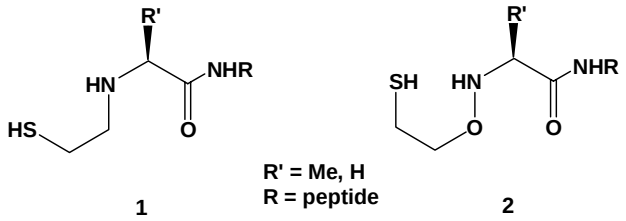


Figure 15: early acyl transfer auxiliaries.

The rate-determining-step in these trials was the intramolecular *S* to *N* acyl shift rearrangement. This contrasts with cysteine ligation, where this rearrangement step is rapid and transthioesterification or thiol exchange is the rate-determining-step. The speed of the rearrangement was severely reduced by steric hindrance caused by the presence of even small substituents on the α-carbon of either the *C*-terminal or *N*-terminal amino acid residues on either side of the ligation junction. Indeed in these cases the rearrangement proceeded so slowly that it was even possible to isolate the transthioesterified but unrearranged product. The longer chain of **2** causes the key intramolecular rearrangement step to take place via a six-membered rather than five-membered ring transition state, in addition to introducing undesirable extra flexibility, and appeared to be the reason for the poor performance shown by this auxiliary.

Despite these initial problems, auxiliaries based on both these basic scaffolds have been developed for use in cysteine-free peptide ligation. The 2-mercaptobenzyl auxiliaries

developed by Dawson and co-workers[129] (see **Figure 16**) undergo rearrangement via a six-membered ring transition state. In this case, however, the six-membered ring is constrained or "pre-organised" by the aromatic ring. The methoxy substituents are designed to improve the efficiency of the auxiliary by increasing the electron density of the aromatic ring and hence increase the nucleophilicity of the thiol group (5-methoxy substituent) and the acid lability of the auxiliary to allow removal following ligation (4- and 6-methoxy substituents).

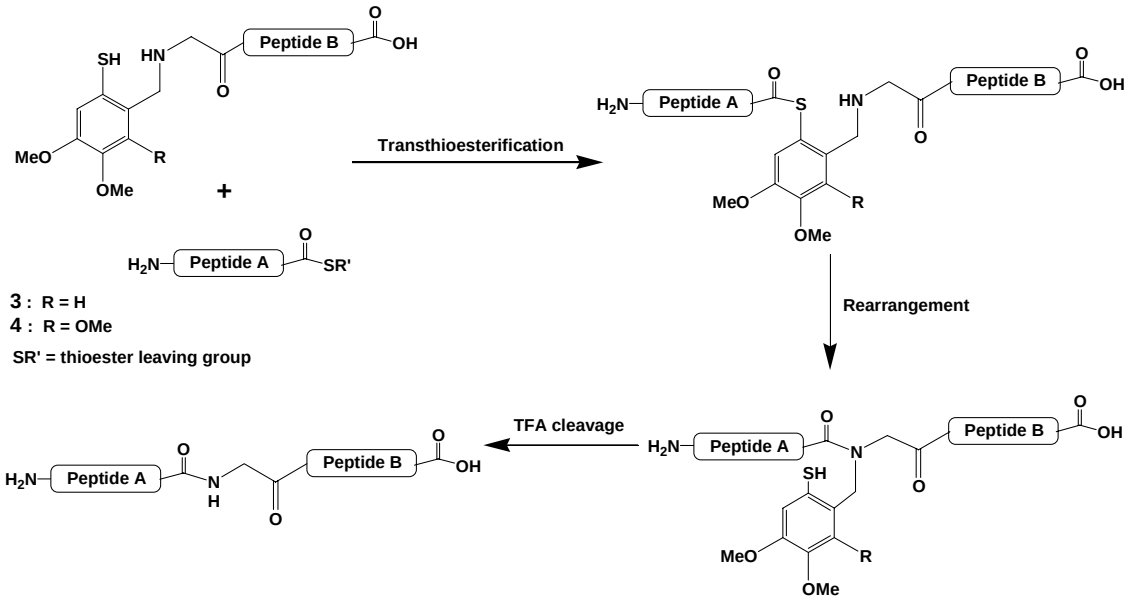


Figure 16: 4,5-dimethoxy-2-mercaptobenzyl (**3**) and 4,5,6-trimethoxy-2-mercaptobenzyl (**4**) auxiliary mediated cysteine-free peptide ligation.

Both 4,5-dimethoxy-2-mercaptobenzyl (Dmb) **3** and 4,5,6-trimethoxy-2-mercaptobenzyl (Tmb) **4** auxiliaries were successfully used in ligations of model peptide segments at Gly-Gly and Lys-Gly junctions in moderate to good yield. The Dmb auxiliary was also used in a successful ligation at the Gly-Ala junction, although

reaction in this case was much slower and less efficient. In contrast to standard conditions for NCL, the presence of thiol additives was found to prevent rather than facilitate ligation. This is likely to be due to the excess thiol intercepting the initial transthioesterified intermediate before the slow six-membered ring rearrangement step can occur. Ligation causes the benzylamine moiety to become acylated and hence labile to acid treatment. The Tmb auxiliary can be removed by treatment with TFA, but the Dmb auxiliary requires the use of HF.

Based on the results of the ligations carried out on model peptide systems, the Tmb auxiliary appears to be an efficient acyl-transfer agent which is tolerant of side chains on the *C*-terminal residue of the thioester. In principle, therefore, the Tmb auxiliary should be useful for cysteine-free peptide ligation at Xxx-Gly junctions, where Xxx represents any amino acid with a non-sterically bulky side chain. It may also be possible to achieve Gly-Ala ligations using this auxiliary. The practical difficulties of associated with HF treatment make the Dmb auxiliary less useful and also render it incompatible with any sensitive modifications present in the peptide or protein, including glycosylation or phosphorylation.

In very recent work by Danishefsky and co-workers,[113] the Tmb auxiliary was used in the assembly of glycopeptides via ligation at Gly-Xxx junctions. Ligation was achieved at a Gly-Ala junction in moderate yield. More impressively, a successful ligation was achieved at the more sterically hindered Gly-Gln junction. Unsurprisingly, the initial attempt at this ligation under standard conditions in phosphate-buffered-saline (PBS) solution was unsuccessful. The addition of DMF cosolvent to the reaction mixture

however, resulted in ligation in an impressive 54 % yield for this difficult junction. A double-ligated product, formed by acylation of the auxiliary with a second equivalent of thioester following rearrangement, was also formed, but this could be reduced to the desired product by treatment with excess thiol. The yield of ligation at the Gly-Ala junction was also improved under these conditions.

In contrast to previous reports[129] however, difficulties were encountered during post-ligation auxiliary removal by treatment with TFA. In addition to the desired cleaved product, formation of a side product of the same mass as the starting material was observed. This was postulated to be the thioester intermediate, arising from a reversal of the $S - N$ acyl shift rearrangement initiated by irreversible protonation of the benzylic amine.[130] To avoid this problem, a two step cleavage protocol was used, involving selective methylation of the auxiliary thiol with methyl *p*-nitrobenzene sulfonate[131] prior to TFA cleavage. This procedure would be unsuitable for peptides containing unprotected cysteine residues, however.

The *N*α-(2-mercaptoethyl) scaffold **1** which was most successful in the initial research into cysteine-free ligation has also been used as basis for the development of removable acyl transfer auxiliaries. The *N*α-(1-phenyl-2-mercaptoethyl) class of auxiliaries developed by Kent and co-workers[132] also uses a substituted aromatic ring to introduce post-ligation lability to the thiol auxiliary, but the rearrangement step in this case takes place via a five-membered ring transition state analogous to cysteine ligation (see **Figure 17**).

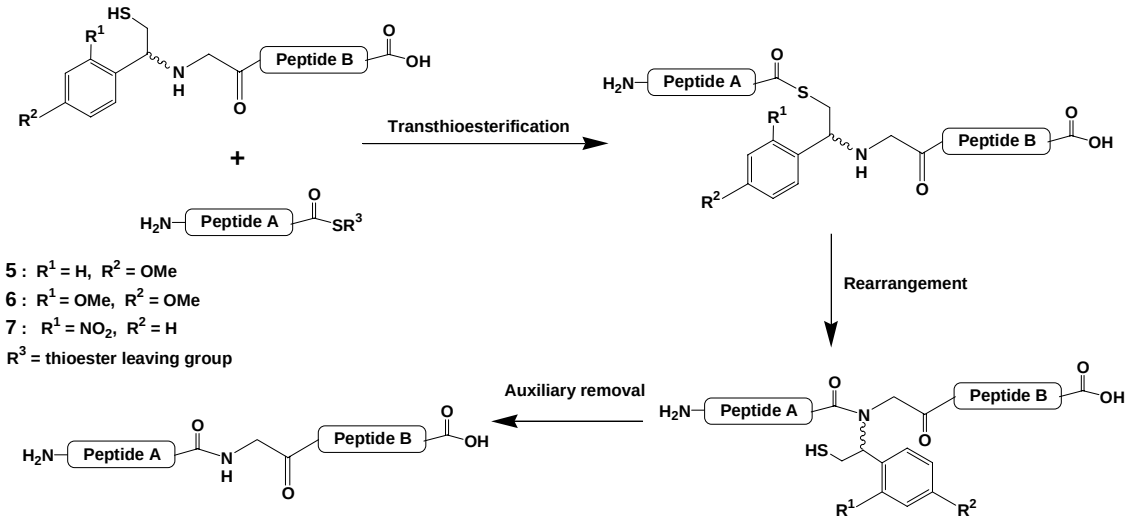


Figure 17: *N*α-(1-phenyl-2-mercaptoethyl) auxiliary mediated cysteine-free peptide ligation.

The *N*α-(1-(4-methoxyphenyl)-2-mercaptoethyl) **5** and *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) **6** auxiliaries were tested in trial ligations with short model peptides for several ligation junctions. Successful ligations were achieved for Gly-Gly, His-Gly and Ala-Gly junctions for both auxiliaries and also at a Lys-Gly junction for the less substituted **5**. This auxiliary has also been successfully used at a Gly-Ala junction in recent studies on the formation of cyclic peptides by cysteine-free ligation.[133] In each case reaction was slower for the more substituted, larger **6**, probably due to the sensitivity of the key rearrangement step to steric hindrance as discussed previously. In contrast to the 2-mercaptobenzyl auxiliaries, thiol additives did not prevent ligation in these trials. The dimethoxy-substituted auxiliary **6** is removable by treatment with TFA treatment following ligation, whereas removal of **5** requires HF, limiting its applicability. **5** has however been used in the total synthesis of the 106-residue protein cytochrome b562 via a ligation at a His-Gly junction.[134] The *N*α-(1-(2,4-

dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** would therefore appear to be applicable to cysteine-free peptide ligation at Xxx-Gly junctions, where the thioester *C*-terminal amino acid residue Xxx has a non-sterically bulky side-chain.

Recent studies by Dawson and co-workers have examined the use of a photolabile *N*α-(1-(2-nitrophenyl)-2-mercaptoethyl) auxiliary **7**.[135] This auxiliary was tested at Gly-Gly and Ala-Gly junctions using a series of short model peptides. Efficient ligations were achieved for both junctions, although as expected reaction rates for the Ala-Gly junction were considerably slower than for the Gly-Gly ligation. Following ligation, the auxiliary was removed by photolysis at 310 nm. Although these reactions were high yielding with minimal formation of side-products, the use of photolytic treatment for larger peptides and proteins is potentially problematic, due to the possibility of oxidative degradation of sensitive residues such as methionine and tryptophan, and other degradative side reactions.

2.3.6 "Traceless" Staudinger Ligation

An alternative approach to cysteine-free peptide ligation is based on the "traceless" Staudinger ligation developed by Bertozzi and co-workers[136,137] and Raines and co-workers.[138,139] In the Staudinger ligation, a chemoselective coupling reaction takes place between an azide and a phosphine to form an aza-ylide intermediate[140] (see **Figure 18**). Hydrolysis of the ylide gives the reduced amine. Alternatively, if an appropriate electrophilic moiety is conjugated to the phosphine, the aza-ylide can be

trapped by intramolecular nucleophilic attack of the charged nitrogen. Hydrolysis of the resulting cyclic intermediate gives an amide-linked product. This reaction has been used to modify surface oligosaccharides in living cells, demonstrating its excellent selectivity.[136]

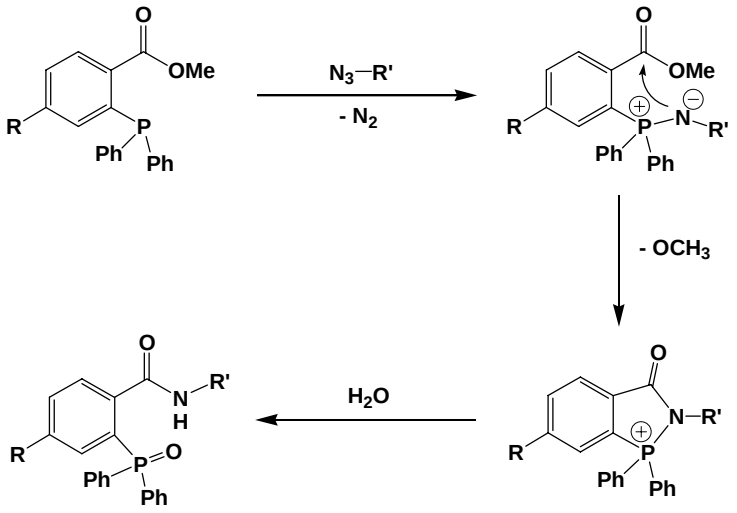


Figure 18: Staudinger ligation to form an amide linkage between **R** and **R'**.

The insertion of a cleavable linker between the phosphine and the species to be coupled (**R** in **Figure 18** above) results in "traceless" Staudinger ligation with the formation of a native amide bond (see **Figure 19**). Several ester and thioester structures have been tested as linkers but the simple phoshinomethanthioester **8** has proved the most effective.[136,139] In common with the acyl transfer mediated ligation, the critical factor in the Staudinger ligation appears to be the intramolecular rearrangement resulting in acylation of the nitrogen atom derived from the azide coupling partner. In the case of phosphinomethanthioester **8** this takes place via a favourable five-membered ring transition state which is constrained by the phenyl substituents and in which the

nucleophilic aza-ylide nitrogen atom is in close proximity to the thioester carbon. **8** is also more stable to hydrolysis than the corresponding ester.[139]

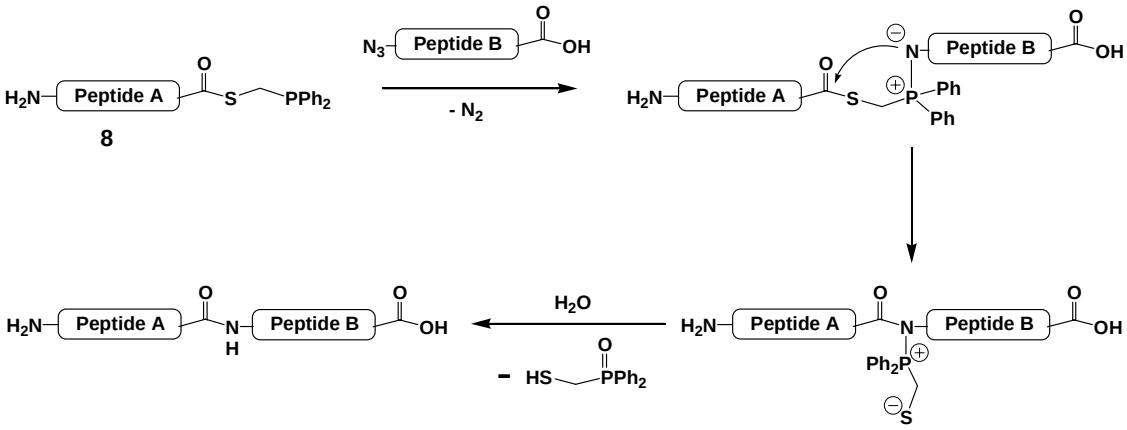


Figure 19: "traceless" Staudinger ligation to form a native amide linkage between two peptide fragments.

Efficient ligations using this technique have been achieved between short model *C*-terminal phosphinothioester and *N*-terminal azide peptide fragments at a range of Xxx-Gly junctions. Reaction conditions are mild and typically involve a mixed organic/aqueous solvent system. In common with the auxiliary-mediated ligation, thioesters with non-sterically bulky *C*-terminal residues give the best results. *C*-terminal glycine, alanine and phenylalanine thioesters have all been ligated to *N*-terminal azidoglycine peptides in good to moderate yields.[139,141] A ligation between a short, protected *C*-terminal glutamic acid thioester and an *N*-terminal azidoglycine peptide has also been reported.[143] In another similarity with auxiliary-mediated ligation, the Staudinger approach is relatively intolerant of sterically bulky *N*-terminal residues. Low yields have been reported for ligations at Gly-Phe and Ala-Phe junctions, although

these still represent an improvement over the results for similar junctions via the auxiliary approach. Unsurprisingly, negligible yields of ligated product were reported for the more sterically demanding Gly-Leu and Ala-Leu junctions.[142]

Recent work has seen the application of this method to protein synthesis and semi-synthesis. Raines and co-workers used a combination of the traceless Staudinger ligation and EPL to synthesise the 124 residue protein RNase A.[143] The *C*-terminal fragment was synthesised as an *N*-terminal azidoglycine peptide on solid phase. Ligation with a protected *N*-terminal cysteine and *C*-terminal phosphinothioester containing peptide fragment was carried out in reasonable yield on solid phase, despite the presence of a sterically bulky glutamic acid side chain at the *C*-terminus. Following concomitant side chain deprotection and release from the resin, the resulting *N*-terminal cysteine peptide was used in EPL with a bacterially-derived thioester to form the full length protein.

Wong and co-workers have recently applied the traceless Staudinger ligation to the synthesis of short model glycopeptides, and have shown that the technique is compatible with glycosylated peptides, provided the glycosylation site is not adjacent to the ligation junction.[141] Only β-linked peracetylated Galactose monosaccharides were used in these trials however, rather than the natural unprotected α-linked GalNac and larger GalNAc derived glycans. A Staudinger ligation between two peptide fragments has therefore yet to be used to synthesise a glycopeptide or glycoprotein displaying natural saccharides.

In the context of applying Staudinger ligation to protein semi-synthesis, bacterially-derived phosphinothioesters are available by the treatment of recombinant intein fusion proteins with a phosphinothiol, in the same manner as described for EPL.[144] Wong and co-workers have also recently developed methods for the modification of unprotected recombinant proteins with *N*-terminal azidoglucose dipeptides to form recombinant *N*-terminal azidoglycine protein fragments.[141] Protease catalysed condensation, using the enzyme subtilisin Carlsberg, was used to couple an *N*-terminal azidoglucose dipeptide trifluoroethyl ester to an unprotected recombinant peptide. The resulting *N*-terminal azidoglycine protein fragment was then ligated to a short model *C*-terminal phosphinothioester peptide glycosylated with β-Gal$(OAc)_4$ to give a semi-synthetic glycopeptide mimic.

"Traceless" Staudinger ligation may offer an attractive alternative to auxiliary-mediated ligation for cysteine-free peptide ligations. It offers similar tolerance of sterically bulky residues and glycosylation around the ligation junction, with some results suggesting that the key rearrangement step may in fact be less inhibited by steric bulk than is the case for auxiliary-mediated ligation.[143] The technique has also been successfully applied to protein semi-synthesis. Another attractive feature is the cleavage of the phosphinothiol group under the ligation conditions, instead of by acidic or photolytic cleavage in an additional step. The assembly of native glycopeptides or glycoproteins has not been demonstrated however, and as yet only a single native protein has been assembled by this approach. The method is also somewhat less flexible than the auxiliary approach, where the auxiliary thiol can be temporarily protected during an initial ligation, then unmasked at the appropriate time for use in a subsequent ligation.

2.3.7 Peptide Ligation via Decarboxylative Condensation

Very recently, the application of a different chemoselective reaction to peptide ligation has been reported by Bode and co-workers.[145] The approach is based upon reaction between an *N*-hydroxylamine and the ketone of an α-ketoacid. This results in the formation of an unstable hemiaminal intermediate which decomposes on mild heating, with the loss of carbon dioxide and water, to give an amide (see **Figure 20**).

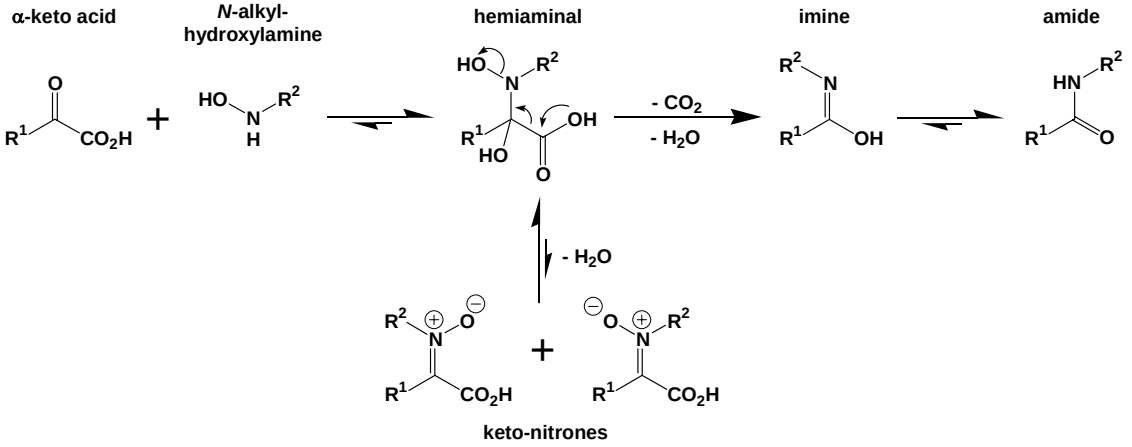


Figure 20: decarboxylative condensation reaction between an *N*-alkylhydroxylamine and an α-ketoacid.

This decarboxylative condensation was used to form native amide linkages between *N*-terminal hydroxylamine-[146,147] and *C*-terminal α-ketoacid-[148] peptide fragments. Successful couplings were reported between short model peptide fragments under mild heating in DMF, aqueous buffers and mixed solvent systems. Both protected and unprotected peptides were used and no further reagents were required. Couplings were complete in 24 hours in generally good yields for all the couplings examined, including

those involving sterically crowded junctions such as Pro-Ala, Val-Gly and Phe-Ala (see **Figure 21**).

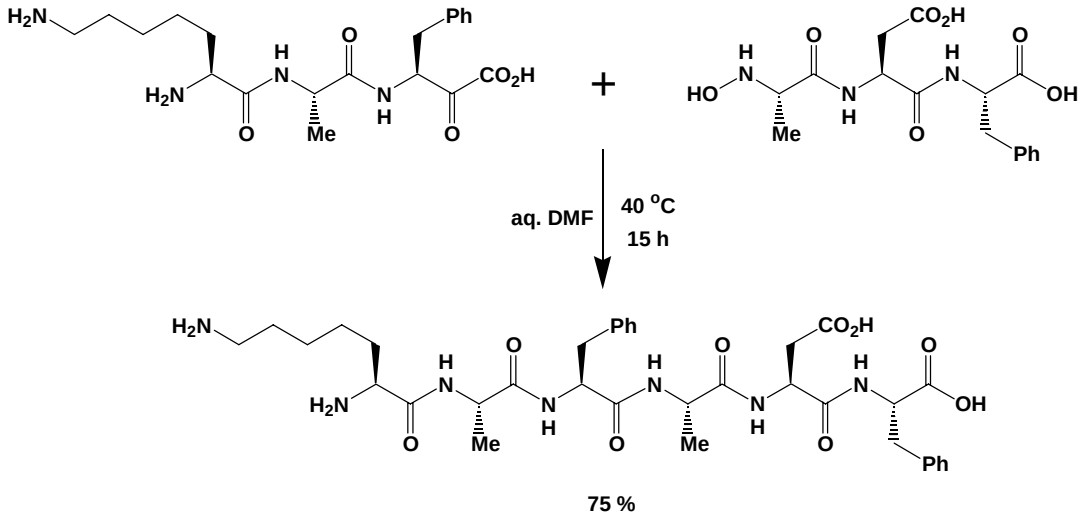


Figure 21: peptide ligation at Phe-Ala junction via decarboxylative condensation.

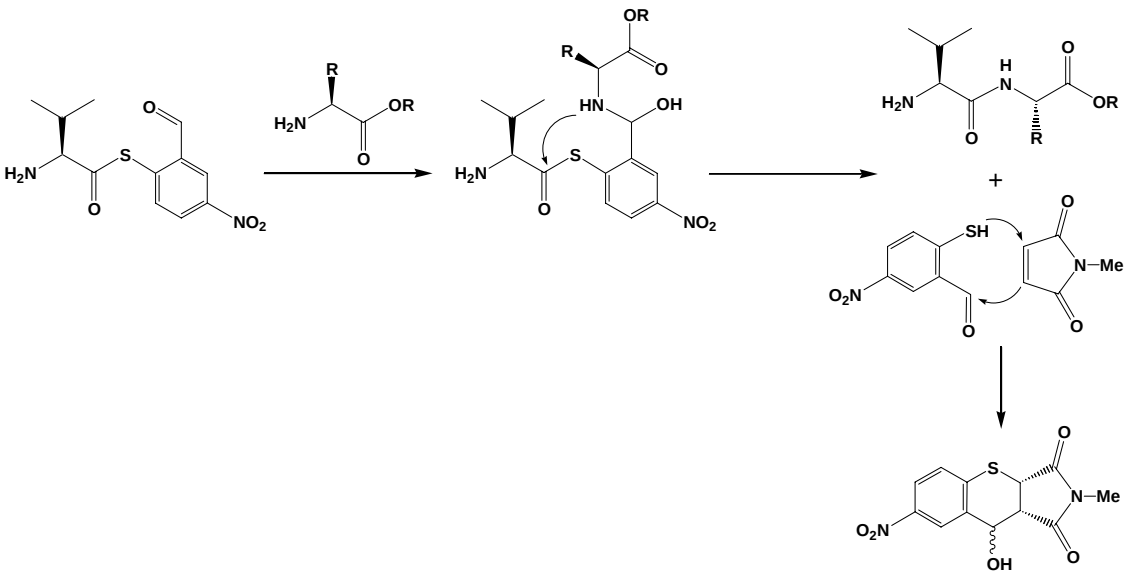
The approach has also been applied to the synthesis of β-oligopeptides via decarboxylative condensation of α-ketoacids with isoxazolidine acetals.[149] Although only preliminary details of this technique have so far been reported and details of the mechanism remain to be elucidated, the mildness of the reagent-free coupling conditions and the apparent tolerance for sterically congested junctions make this new technique a potentially valuable tool for peptide ligation.

2.3.8 *C*-Terminal Activation Methods

A different approach to protein assembly by the coupling of peptide fragments is taken in the "thioester method" developed by Hojo and co-workers.[150] In this technique, the *C*-terminal thioester of a peptide fragment is transformed into an active ester in the presence of a silver (I) salt catalyst, typically by treatment with 3,4-dihydro-3-hydroxy-4-oxo-1,2,3-benzotriazine (HOOBt) and base in organic solvent. A standard peptide coupling then occurs with the *N*-terminal primary amine of the second peptide fragment. This method has been used to assemble several large peptides and also a model mucin-type glycopeptide displaying multiple α-GalNAc saccharides.[151]

The method is not fully chemoselective and any lysine or cysteine side chains present in either peptide fragment must be protected prior to coupling. This necessitates extra protection/deprotection and purification steps in the peptide assembly process. Some cleavage of the acetamido groups typically used to protect the cysteine side chain, and consequent side product formation, has also been reported after prolonged exposure to the coupling conditions in slow reactions.[150] In contrast to the cysteine-free ligation methods previously described however, couplings by the "thioester method" are not inhibited by sterically bulky side-chains on the *N*-terminal amino acid and efficient couplings have been reported for a range of Gly-Xxx junctions. The method has been rarely used for the coupling of non-Gly *C*-terminal thioesters, however, due to epimerization of the *C*-terminal residue under the coupling conditions.[150,151] Despite these drawbacks, the thioester method may present an attractive complementatry strategy to NCL and cysteine-free peptide ligation for the coupling of small synthetic

peptide and glycopeptide fragments, as it allows the use of a different range of potential ligation junctions.



Scheme 10: peptide coupling using the 2-formyl-4-nitrophenyl thioester auxiliary and trapping of byproduct with *N*-methylmaleimide.

In a converse approach to that used in the auxiliary-mediated and Staudinger peptide ligations previously discussed, Ito and co-workers incorporated a formyl-substituted auxiliary into the *C*-terminal thioester peptide fragment (see **Scheme 10**).[152] Selective reaction with the primary amine at the *N*-terminus of a second peptide fragment in the presence of base was followed by an intramolecular *S* to *N* acyl shift rearrangement via a constrained six-membered ring transition state similar to that undergone in the case of the 2-mercaptobenzyl auxiliary. The rearrangement gave the desired amide bond and the 2-formyl-4-nitrobenzenethiol byproduct, which can be trapped by intramolecular aldol cyclization with *N*-methylmaleimide as shown, providing an additional driving

force for the reaction. The 2-formyl-4-nitrophenyl thioester auxiliary was tested for a range of coupling junctions and proved remarkably tolerant of steric bulk on both the *N*-terminal and *C*-terminal amino acid residues. Couplings between the Fmoc-protected 2-formyl-4-nitrophenyl thioester of the β-branched valine residue and a range of amino acids, including sterically bulky residues such as a second valine residue and tyrosine, were achieved in good yields. The reaction was also shown to be compatible with amino acids glycosylated with unprotected saccharides and with the presence of aqueous buffer systems.

Unfortunately, despite the range of junctions examined, only the couplings of single amino acid residues were studied. The utility of the method for the coupling of peptide fragments has therefore yet to be demonstrated. The drawbacks of a *C*-terminal thioester activation approach, such as the necessity for protection of lysine and cysteine residues, as described for the "thioester method" also apply to this coupling technique.

A novel approach to the synthesis of mucin-type glycopeptide mimics has been used by Nishimura and co-workers.[153] Glycosylated tripeptides consisting of threonine glycosylated with the Core 1 disaccharide or a simple Galβ(1→3)Gal disaccharide analogure, flanked on either side by an alanine residue, were initially synthesised. Activation of the *C*-terminal carboxylate group with diphenylphosphoryl azide (DPPA) led to a polymerisation reaction forming mucin-like glycopeptide mimics made up of chains of 8-15 tripeptide units, the "anti-freeze" properties of which were subsequently studied. The polymerisation conditions were compatible with the presence of unprotected saccharides. The limitations on the size of repeating units which could be

used in this method are unclear. The incompatibility of the polymerisation conditions with carboxylate or primary amine side chains also limit the utility of the method in the context of preparing natural glycopeptides.

2.3.9 Protease-Catalysed Peptide Ligation

Protease enzymes function naturally by hydrolysing peptide bonds. They can be used under thermodynamic and/or kinetic control to catalyse the formation of the peptide bond, however.[39,53,154] Under thermodynamic control, the reaction conditions are tuned to favour formation of the amidolysis rather than hydrolysis product, typically by such means as the use of organic solvents to supress ionization of starting materials, or the use of an excess of one of the coupling partners. Under kinetic control, the carboxylate acyl donor is typically activated as an ester to favour the formation of the acyl-enzyme intermediate.

The subtilisin proteases have been widely used in peptide coupling reactions, and engineered subtilisin variants with reduced hydrolytic activity and increased amidolysis activity have been produced, some of which are commercially available.[155,156] Subtilisin-catalysed peptide ligation has been used to assemble several polypeptides, including *N*-linked and *O*-linked glycopeptides.[156-158] These ligations demonstrate excellent regio- and stereo-specificity and take place in the presence of unprotected amino acid side chains and are quite tolerant of different acyl donor and acyl acceptor substrates. The process is particularly useful for the modification or coupling of short

peptide or glycopeptide fragments, as demonstated by Wong and co-workers in the coupling of an *N*-terminal azidoglycine dipeptide ester with a short recombinant polypeptide.[141]

Despite their relatively broad substrate tolerance, the range of acyl donor and acceptor substrates which can be used in subtilisin-catalysed ligation is still inherently limited and a screening and optimisation process must often be used to determine the feasibility of ligation between particular substrates. Subtilisin variants with appropriate tolerance for the desired substrates will also not necessarily be available. Despite the reduced hydrolytic activity displayed by engineered subtilisin variants, some hydrolysis byproducts are generally still formed in these reactions.[141,157,158] These side reactions are likely to become increasingly problematic with larger peptide fragments, especially those containing sequences naturally hydrolysed by the enzyme used. To prevent self-condensation, the *N*-terminus of the acyl donor, and the *C*-terminus of the acyl acceptor must be protected, typically as an Fmoc carbamate and an amide, respectively. Another limitation is the requirement for a large excess of either the acyl acceptor or donor, to minimise the hydrolysis reaction. This can be problematic if both peptide fragments are precious materials.[141]

2.3.10 Synthesis of thioester peptides

With the exception of the protease-catalysed peptide ligation approach described above and the recently reported decarboxylative condensation technique (see section **2.3.7**), all

the methods previously discussed for the ligation of peptide segments involve a *C*-terminal thioester as one of the coupling partners. Recombinant peptide thioester fragments can be produced by bacterial expression, but for a truly general and versatile assembly approach, access to synthetic thioesters is also necessary. Thioester peptides can be produced by Boc-SPPS, using thioester linkers.[159] As previously discussed, the HF treatment necessary for cleavage of the peptide from the solid phase is not compatible with sensitive modifications present in some peptides, such as glycosylation, as well as being practically inconvenient. The repeated treatments with TFA used to remove the Boc amine protecting group can also cause damage to the glycosidic linkage. On the other hand, the thioester group is sensitive to nucleophiles and so is incompatible with the repeated piperidine treatments used in Fmoc-SPPS. A range of different approaches have been used to allow the synthesis of thioesters by Fmoc-SPPS, from special deprotection protocols to the use of specialised linkers.

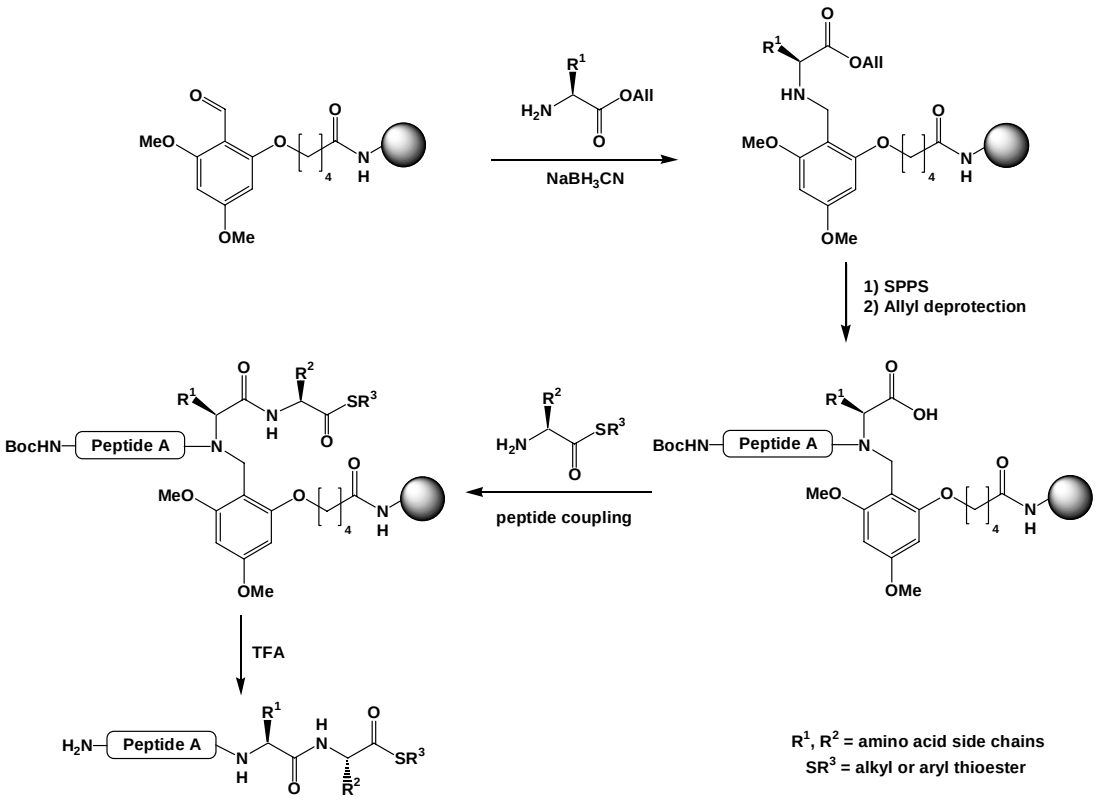
The most straightforward approach to the Fmoc-SPPS of thioesters is the use of deprotection protocols for the removal of the Fmoc group which do not attack thioesters, and therefore allow thioester linkers to be used. Li and co-workers replaced the standard piperidine solution in DMF normally used to remove Fmoc with a solution of 1-methylpyrrolidine, in a 1:1 v/v mixture of NMP and DMSO with the presence of small amounts of hexamethyleneimine and 1-hydroxybenzotriazole (HOBt).[160] Fmoc-SPPS with this deprotection protocol allowed the direct sythesis of short thioester peptides in acceptable yields. Aminolysis of the thioester was still observed however and accounted for a decrease in yield of up to 15 % in model studies. The hindered alkyl thioesters which were found to be the most stable during SPPS in this report were

not tested in subsequent ligations, for which they may be poor substrates.[107] Despite these drawbacks, this deprotection protocol has been used in the synthesis of a partially protected, 24-residue peptide alkyl thioester glycosylated with the *N*-linked core pentasaccharide, which was subsequently used in a silver catalysed thioester ligation.[64] Clippingdale and co-workers developed an alternative deprotection protocol using the non-nucleophilic base 1,8-diazabicylco[5.4.0]undecane-7 (DBU) in the presence of HOBt.[161] Acceptable yields of short peptide thioesters were reported using this system. Aspartamide formation during peptide synthesis was a significant problem with this technique, however.

The most commonly used approaches to peptide thioesters avoid the use of thioester linkers. Instead, the thioester is formed following peptide synthesis. Peptides synthesised on common ester-linker resins, such as Wang or PAM resins, can be treated with alkylaluminium thiolate, formed *in situ* from alkylaluminium chloride and ethanethiol, to form thioesters with concomitant cleavage from the solid phase.[162] Subsequent side chain deprotection by TFA treatment afforded unprotected thioester peptides in moderate to good yields for the short peptides studied. Epimerisation of the *C*-terminal residue was a problem with this method, however, as was unwanted thioester formation at aspartate and glutamate side chains in the peptide.

In the "backbone amide linker" (BAL) approach, the resin linker is attached at the amine nitrogen of the *C*-terminal amino acid residue instead of at its carboxylate group, which is instead protected, typically as an allyl ester. Following peptide chain elongation, the allyl protecting group can be selectively removed and the resulting free

C-terminal carboxylate can be coupled to an amino acid thioester under standard peptide coupling conditions. Acid cleavage from the solid phase and concomitant side chain deprotection affords the unprotected peptide thioester (see **Scheme 11**).



Scheme 11: synthesis of thioester peptides via backbone amide linker approach.

Good yields and minimal racemisation of the penultimate amino acid residue during the final amino acid thioester coupling step were reported for a series of short peptide thioesters.[163] The drawbacks of this strategy derive from the difficulty of the first peptide coupling, between the secondary amine of the resin-linked *C*-terminal residue and the next amino acid in the peptide sequence. The symmetrical anhydrides of the appropriate amino acids, in a mixed solvent system, gave the best results in this

coupling. Different amine protection was also required for the second amino acid, as a large amount of diketopiperazine formation was observed due to intramolecular attack of the *N*-terminal primary amine at the allyl ester under Fmoc removal conditions. The acid labile *N*-2-(3,5-dimethoxyphenyl)propyl[2]oxycarbonyl (Ddz) protecting group was therefore used, although carefully controlled conditions were necessary for its removal to avoid cleavage of the resin linker.

One solution to the problem of diketopiperazine formation is the attachment of the linker at the backbone amine of the *C*-terminal thioester, which is protected as the trithioortho ester, rather than the penultimate residue, as reported by Jensen and co-workers.[164] Following completion of the peptide, the thioester is formed concomitantly with cleavage from the solid phase and side chain deprotection by treatment with TFA. This method is only practical for the synthesis of *C*-terminal glycine thioesters, however, as racemic thioesters will be formed for peptides with chiral *C*-terminal residues.

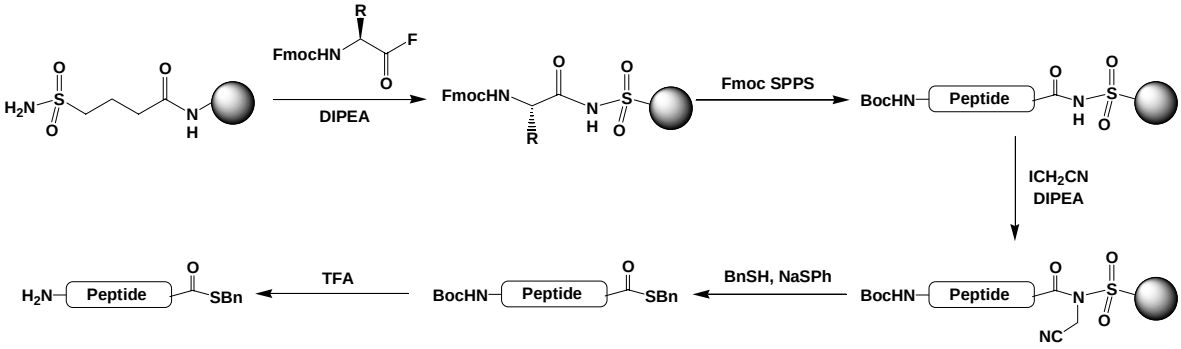
The use of highly acid labile linkers, such as the carboxy-trityl[165,166] or 2-chlorotrityl[167] linkers, allows peptides to be cleaved from the solid phase under mildly acidic conditions, such as treatment with acetic acid[165,167] or 5 % v/v TFA[166] solutions, which leave the side-chain protecting groups intact. Activation of the *C*-terminus of the resulting peptide with a smaller excess of the coupling reagents benzotriazole-1-yl-oxy-tris-pyrrolidinophosphonium hexafluorophosphate (PyBOP)[167] or HBTU[165,166] than is typically used in peptide coupling, and coupling of a thiol, followed by side chain deprotection with TFA, affords the corresponding unprotected thioester. Good yields

and low racemisation have been reported for the synthesis of peptide and glycopeptide[166] thioesters via this route. The practical incovenience caused by the requirement for a final solution phase coupling and subsequent purification step may be considered a drawback to this approach, however.

The most general and widely used method for the Fmoc SPPS of thioesters is based on a modification of Kenner's sulfonamide "safety-catch" strategy.[168-170] In this approach, the *C*-terminal amino acid residue of the target peptide is coupled to the solid phase via an alkanesulfonamide linker, creating an *N*-acyl sulfonamide linkage. This linkage is acid-stable and is also resistant to basic hydrolysis, as the acyl sulfonamide nitrogen is deprotonated under basic conditions and the linkage is thus protected from nucleophilic attack. Following peptide completion, the acyl sulfonamide is activated by substitution with an electrophilic reagent such as diazomethane or iodoacetonitrile in the presence of base. Alkylation of the sulfonamide, particularly with an electron-withdrawing substituent, renders it susceptible to nucleophilic attack, which cleaves the peptide from the linker. The use of nucleophilic thiols such as thiophenol and benzyl mercaptan in this step results in formation of the corresponding *S*-phenyl or *S*-benzyl thioesters, respectively (see **Scheme 12**). Side chain deprotection following this nucleophilic cleavage from the resin affords the unprotected peptide thioester.

Peptide thioesters have been synthesised in good yields with low reported *C*-terminal racemization by this method.[169] The range of glycopeptides and glycopeptide mimics, containing glycans of up to heptasaccharide size, which have also been synthesised

using this technique demonstrates the utility of the method for the production of glycosylated peptide thioesters.[62,85,108,118]



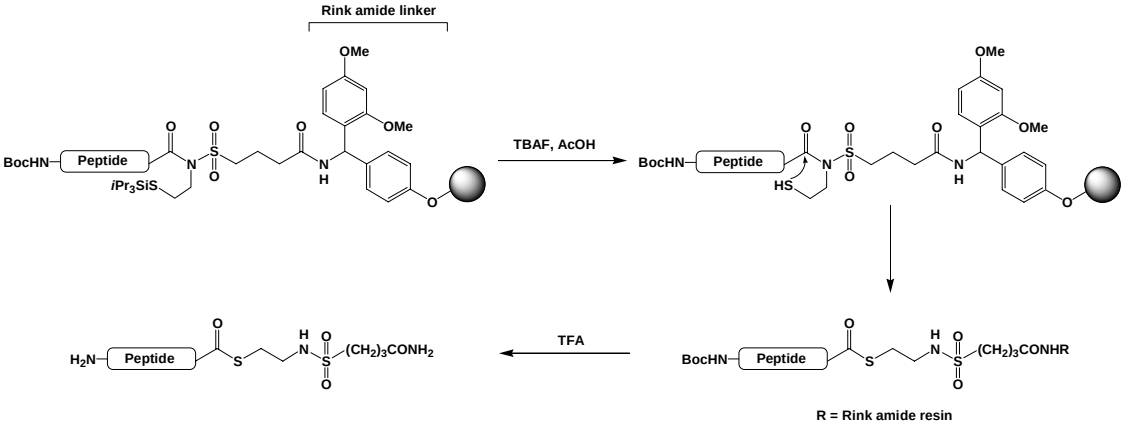
Scheme 12: synthesis of peptide thioesters by Fmoc SPPS on alkanesulfonamide "safety-catch" resin.

Drawbacks of the approach include inefficient coupling of the first amino acid, although this can be overcome by the use of acid fluorides[171] or the PyBOP coupling reagent,[170] in the presence of Hünig's base, in appropriate solvent systems. The two-step activation-cleavage procedure can also be problematic. Of the electrophilic reagents commonly used to activate the sulfonamide, diazomethane and trimethylsilyldiazomethane are not always sufficiently activating, resulting in an incomplete cleavage reaction. Alkylation with iodoacetonitrile results in a more electron-withdrawing, more activating *N*-alkyl substituent but in this case the alkylation reaction itself can be inefficient. The proton of the acyl sulfonamide nitrogen atom is sufficiently acidic to participate in a Mitsunobu reaction, and this approach has also been used for alkylation/activation.[172]

One of the main drawbacks of the "safety-catch" method is the difficulty associated with monitoring reactions, as the two-step activation and cleavage procedure must be used to obtain samples for analysis. Analysis of reaction progress on safety-catch resins is therefore considerably more time-consuming than for standard ester-linker resins. To allow easier monitoring of synthetic progress, particularly the key loading of the first amino acid and electrophilic activation steps, Unverzagt and co-workers recently applied a "double-linker" concept to the synthesis of a thioester on safety-catch resin.[62] A phenylalanine residue was coupled to amino-PEGA resin derivitised with a Rink amide linker. Subsequent coupling of 3-carboxypropanesulfonamide gave an alkanesulfonamide linker which could be cleaved at the phenylalanine residue by treatment with TFA to allow easy monitoring of synthetic progress. The "double-linker" safety catch resin was used to synthesise a 38 residue peptide thioester corresponding to a fragment of RNase B which was glycosylated with an *N*-linked complex-type heptasaccharide, in good yield.

Other recent innovations in safety-catch thioester synthesis have utilised intramolecular reactions for the formation of the thioester. Melnyk and co-workers alkylated several model peptides on safety-catch resin with triisopropylsilyl-protected ethanethiol under Mitsunobu conditions.[173] Fluoride-mediated removal of the triisopropylsilyl protecting group led to an intramolecular *N* to *S* acyl shift via a favourable five-membered ring transition state, in the reverse process to that undergone in native chemical ligation. The peptide thioesters were thus formed on the solid phase. When a Rink amide-safety catch "double-linker" approach similar to that described above was employed, TFA cleavage from the resin with concomitant side chain deprotection gave the unprotected

peptide thioesters in good yield (see **Scheme 13**). Although bulky *C*-terminal residues have been found to inhibit such rearrangements in ligations, efficient thioester formation was reported for a thioester with the β-branched valine residue at the *C*-terminus, suggesting that this approach could be used to synthesise thioesters with a wide range of *C*-terminal residues.



Scheme 13: thioester synthesis via intramolecular *N* to *S* acyl shift.

A similar intramolecular *N* to *S* rearrangement approach has been used by Aimoto and coworkers.[174] In this case the rearrangement was mediated by the 4,5-dimethoxy-2-mercaptobenzyl auxiliary **3** discussed earlier (see section **2.3.5**). The auxiliary, with its thiol masked by the trityl protecting group, was conjugated to Fmoc-alanine and was coupled to PAM resin via an intervening alanine spacer. Following peptide elongation, TFA treatment cleaved the peptide side chain and auxiliary thiol protecting groups, whilst leaving the PAM linker intact. The unmasking of the auxiliary thiol initiated an *N* to *S* acyl shift via a constrained six-membered ring transition state to give the resin-bound thioester, which could be released from the solid phase by transthioesterification

with MESNa. In common with the use of the auxiliary in cysteine-free ligation, the steric bulk of *C*-terminal residue has a profound influence on the efficiency of the rearrangement. The yield for a model leucine *C*-terminal thioester was much lower (5 %) than that of the corresponding glycine thioester (31 %). Interestingly, poorer results were obtained when the auxiliary was conjugated to glycine instead of alanine. It was postulated that the steric bulk of the alanine side chain was actually facilitating the reaction by slowing down the reverse *S* to *N* acyl shift. Yields for the thioesters synthesised in this methods were generally lower than those of the above safety-catch approaches, probably due to slow *N* to *S* rearrangement and also to the difficulties associated with the use of secondary amino acids (the auxiliary-alanine conjugate, in this case) in peptide couplings.

In one of the most interesting recent developments in the field of native chemical ligation, Danishefsky and co-workers utilised another intramolecular rearrangement to prepare thioesters for NCL *in situ* from phenolic esters.[66,112,175] Model peptides were prepared by standard Fmoc SPPS using a resin linker system which allowed cleavage from the solid phase whilst leaving side-chain protection intact. A solution phase peptide coupling was then carried out between the protected peptide and phenylalanine phenolic ester, equipped with an *ortho*-disulfide group, to afford the unprotected peptide phenolic ester following side chain deprotection. Unmasking of the *ortho*-thiol group occurred under the reducing conditions used in standard NCL, initiating a dynamic equilibrium between the peptide phenolic ester and thioester. As the thioester was intercepted by the *N*-terminal cysteine peptide fragment or by the thiol additives in the

reaction mixture however, the equilibrium altered to fully convert the phenolic ester to the thioester and allow the ligation to go to completion (see **Figure 22**).

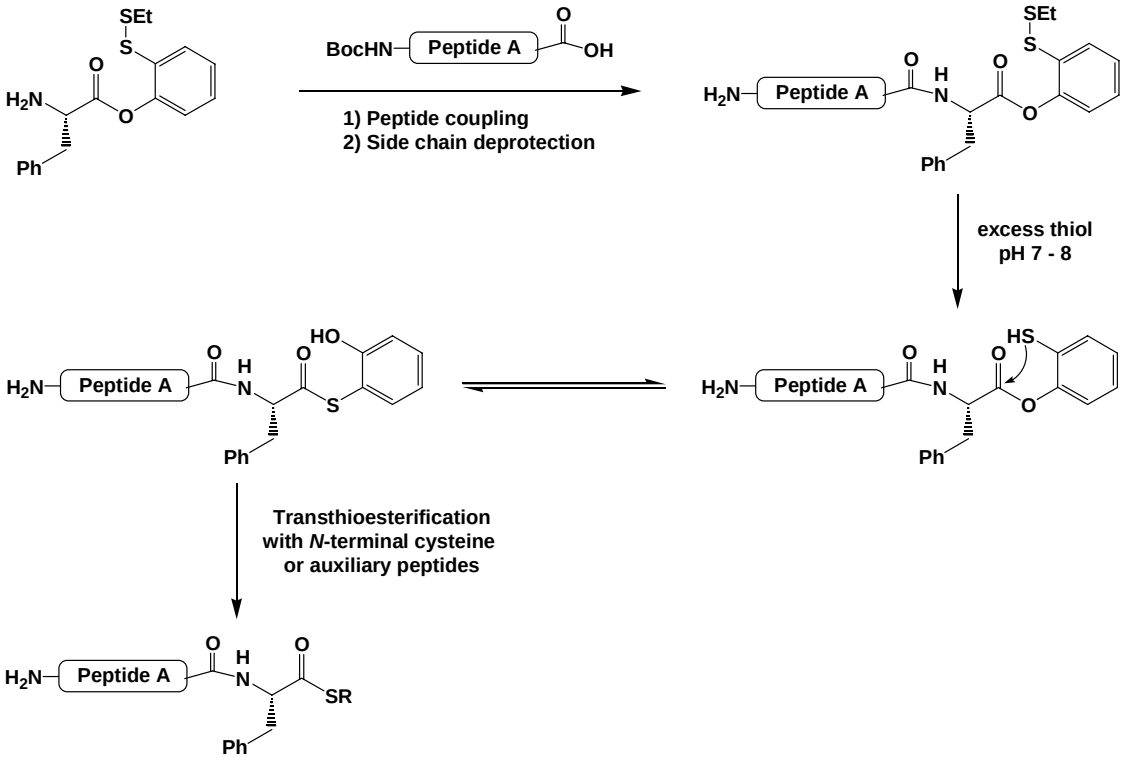
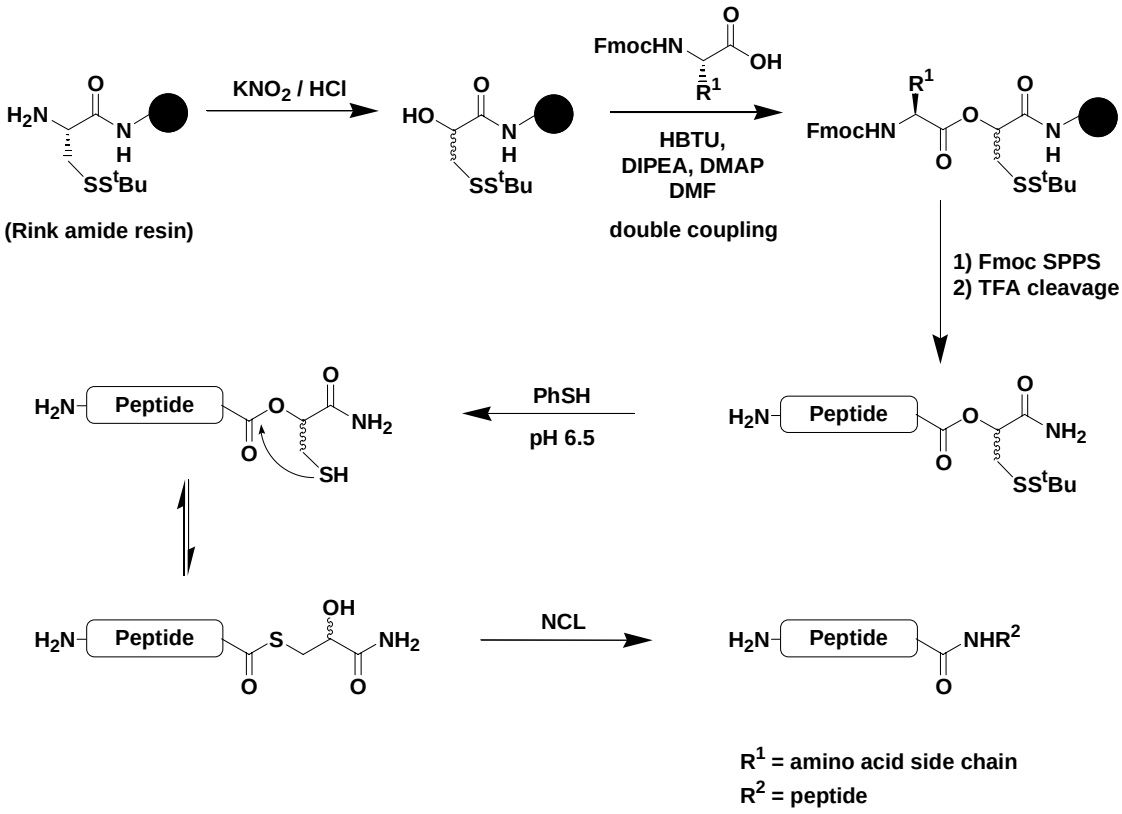


Figure 22: *in situ* thioester formation from peptide phenolic ester

This approach has been used to assemble a range of peptides and glycopeptides containing large saccharide structures via NCL. Yields were generally high, with minimal racemisation observed, although the necessity for partially-protected peptides and the solution phase coupling of the phenylalanine phenolic ester can be considered drawbacks to the method, as previously discussed in relation to the use of highly acid labile linkers. The extension of this technique to the assembly of glycopeptides by cysteine-free ligation via the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4** has very recently been reported,[113] as discussed in section **2.3.5**. A particularly noteworthy

aspect of this report is the success of the *in situ* thioester formation despite the absence of excess thiol additives in the ligation conditions used.



Scheme 14: *in situ* thioester formation from *C*-oxy(2-mercapto-1-carboxyamide)-ethyl ester and subsequent use in NCL.

A related approach, which avoids the necessity for a solution phase coupling, involves the *in situ* rearrangement of peptide *C*-oxy(2-mercapto-1-carboxyamide)-ethyl esters.[176] These can be synthesised on Rink amide resins as shown in **Scheme 14**. Following release from the resin, the cysteine thiol group can be unmasked under standard ligation conditions in the presence of excess thiol, allowing an *O* – *S* acyl shift to take place and resulting in the formation of the thioester. Significant formation of the peptide acid

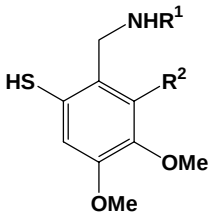
resulting from hydrolysis of the ester and thioester was reported during cleavage from the solid phase and also during the subsequent ligation however, and this problem can be considered a major drawback to this method.

3. Project Strategy

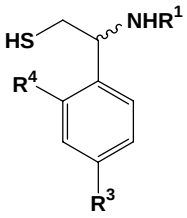
One of the general aims of the research in our group is the development of effective and general methods for the assembly of glycoproteins. As discussed previously, chemoselective reactions are of particular utility in this field due to the sensitive and densely functionalised nature of glycoproteins. Peptide ligations such as NCL and EPL are especially valuable techniques for glycoprotein assembly and have been widely used for this purpose. The requirement for an *N*-terminal cysteine peptide fragment limits the generality and versatility of these methods, however. We were interested, therefore, in investigating the application of cysteine-free peptide ligation to the assembly of glycoproteins.

Of the methods previously discussed, post-ligation reduction of cysteine or selenocysteine is likely to be inappropriate for the assembly of large proteins due to the potential for damage to certain amino acid side chains. Both the "thioester" and formyl auxiliary *C*-terminal activation approaches are not fully chemoselective, as they require partially protected peptide fragments. This drawback also makes them inapplicable to EPL. Protease-catalysed peptide ligations also require partially protected peptide fragments, in addition to their lack of generality and drawbacks for large proteins.

At the outset of the project, the Staudinger ligation had only been reported for the coupling of short, partially-protected peptide fragments at Xxx-Gly junctions and its use had not appeared in the literature subsequent to the papers describing its development. The difficulty of producing recombinant *N*-terminal azidoglycine peptide fragments also appeared to limit its utility in EPL. Furthermore, sequential couplings using this method appeared potentially problematic. We therefore chose to investigate the thiol-auxiliary approach, which seemed to be the most potentially general and versatile approach to cysteine-free peptide ligation.



3 : $R^2 = H$
4 : $R^2 = OMe$



5 : $R^3 = OMe$, $R^4 = H$
6 : $R^3 = OMe$, $R^4 = OMe$
7 : $R^3 = H$, $R^4 = NO_2$

R^1 = peptide

Ligation Junction	Auxiliary used	References
His-Gly	**5, 6**	132, 134
Lys-Gly	**3, 4, 5**	129, 132
Gly-Gly	**All**	129 - 135
Ala-Gly	**5, 6, 7**	132, 135
Gly-Ala	**3, 4, 5**	113, 133
Gly-Gln	**4**	113

Table 1: ligation junctions reported for thiol auxiliaries **3 - 7**.

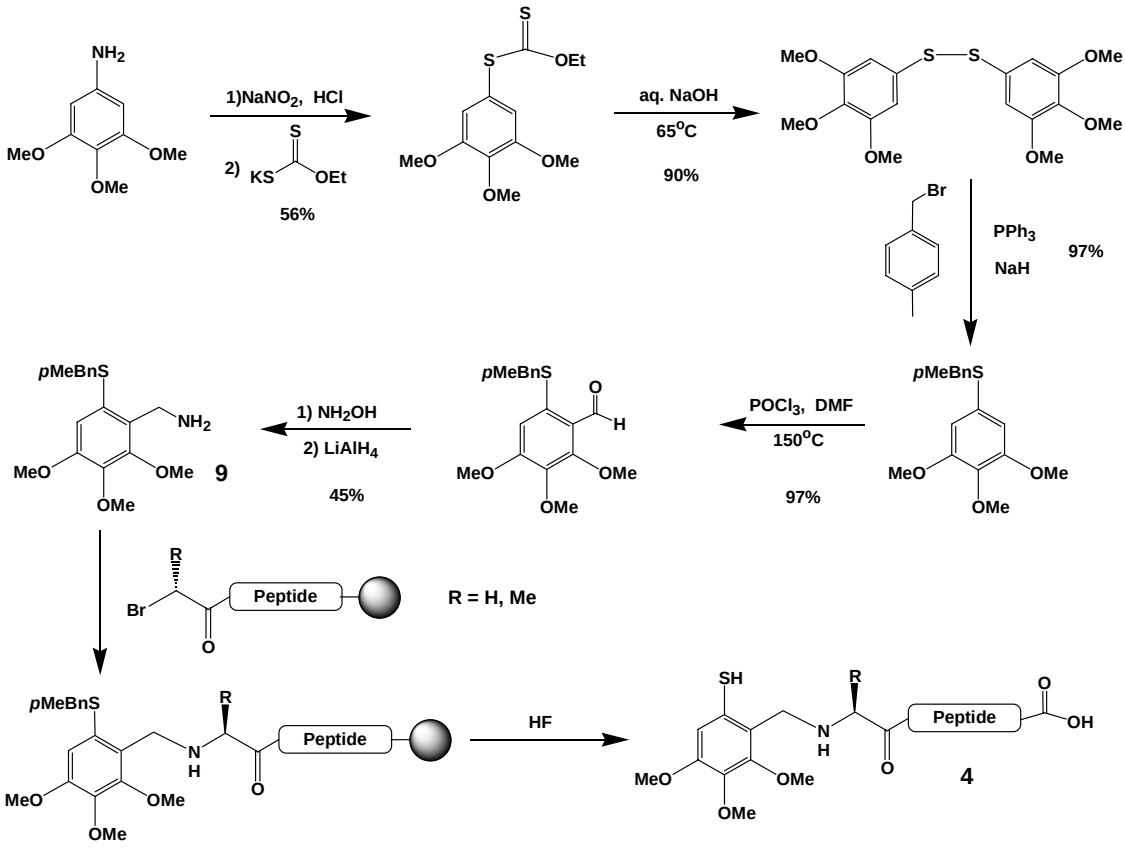
Several successful ligations had been reported using both the 2-mercaptobenzyl (**3** and **4**) and *N*α-(1-phenyl-2-mercaptoethyl) (**5** and **6**) classes of auxiliary, a number of which had gone beyond the use of very small test peptide fragments.[129] Indeed an *N*α-(1-(4-

methoxyphenyl)-2-mercaptoethyl) **5** auxiliary-mediated ligation had been successfully used in the total synthesis of the protein cytochrome b562.[134] These results are summarised in **Table 1**. It should be noted however, that the reporting of isolated yields for both ligations and auxiliary removal is sporadic throughout these reports, especially for the ligation of larger peptide fragments.

We aimed to examine the scope and limitations of cysteine-free peptide ligation using both classes of auxiliary, including the cleavage reactions, using TFA-cleavable auxiliaries **4** and **6**. We planned to gather synthetically relevant results by testing these auxiliaries for the ligation of larger peptide fragments than had previously been reported. We were especially interested in investigating ligations at Ser-Gly and Leu-Gly junctions, as these are present at appropriate sites in the GlyCAM-1 peptide sequence to allow assembly by a cysteine-free ligation strategy.[118] We also intended to test the applicability of these auxiliaries to the ligation of glycopeptide fragments. Auxiliaries **3** and **5** were not investigated, as their removal requires treatment with HF, which renders them incompatible with glycosylated peptides. Towards the end of the project, the photocleavable *N*α-(1-(2-nitrophenyl)-2-mercaptoethyl) **7** auxiliary was also reported for ligations at Ala-Gly and Gly-Gly junctions.[135]

The reported syntheses of auxiliaries **4** and **6** and their conjugation to the *N*-terminus of peptides presented some problems for their application to the assembly of glycopeptides, however. In the case of auxiliary **4**, the protected 4,5,6-trimethoxy-2-mercaptobenzylamine **9** was assembled by a prolonged five-step synthesis which

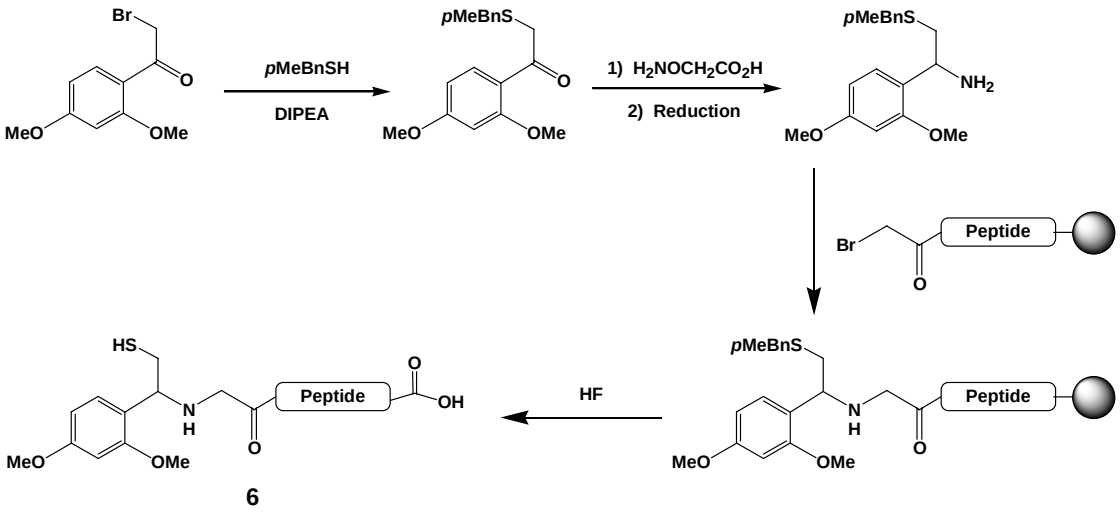
contained two low yielding steps, one of which was the reductive amination of a benzaldehyde, a normally high-yield reaction (see **Scheme 15**).



Scheme 15: literature synthesis of 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4**.[129]

The lengthy "sub-monomer" approach was then used to introduce the auxiliary amine to the *N*-terminus of a model peptide. Firstly, an α-bromoacetyl-anhydride (prepared from the corresponding α-bromo-acid) was coupled to the *N*-terminus of the model peptide on the solid phase. Displacement by **9** gave the desired auxiliary-peptide conjugate after a further three steps. Comcomitant peptide side chain deprotection and release from the solid phase, followed by deprotection of the auxiliary thiol, gave the fully unprotected auxiliary-peptide conjugate ready for use in ligation. The use of the *para*-

methylbenzyl protecting group for the thiol necessitated the use of HF treatment for its removal and hence this synthesis is incompatible with glycosylated or otherwise sensitive peptides in addition to being lengthy and inefficient.



Scheme 16: literature synthesis of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)auxiliary **6**.[132]

The initial syntheses of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** was reported in very little detail (see **Scheme 16**).[132] In common with the synthesis of **4**, however, it involved a two-step amination-reduction procedure and also made use of the *para*-methybenzyl thiol protecting group and hence was incompatible with sensitive peptides. The "sub-monomer" approach was also used for formation of the auxiliary-peptide conjugate. A more efficient synthesis of *N*α-(1-(4-methoxyphenyl)-2-mercaptoethyl) **5** has recently been reported,[133] although this also uses the para-methybenzyl group and so is again not generally applicable. The route to the nitrophenyl auxiliary **7** used a different approach which utilised the TFA-labile Trityl

group to mask the thiol and is therefore more generally applicable, but still involved several steps and "sub-monomer" conjugation to the peptide.[135]

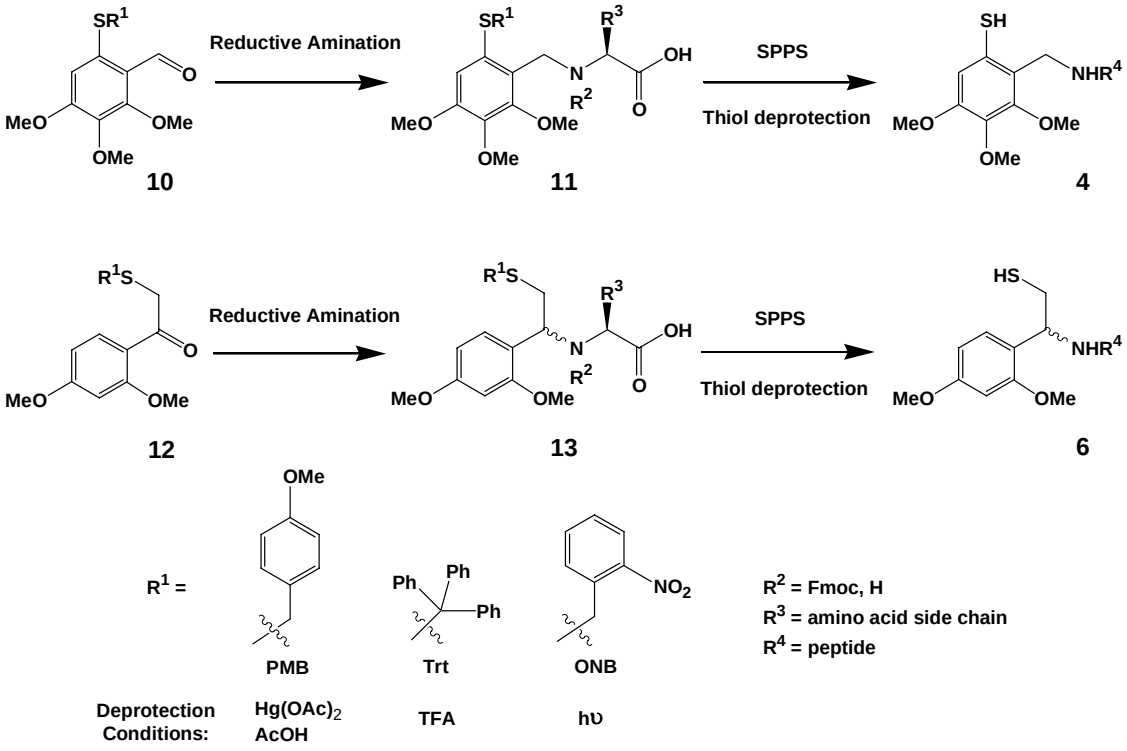


Figure 23: synthetic strategy for auxiliary peptides **4** and **6** via direct reductive amination.

To overcome these problems and allow the application of these auxiliaries to glycopeptide assembly, we aimed to design new short and high yielding synthetic routes to **4** and **6**. We envisaged that this objective could be achieved by a direct reductive amination of the corresponding aldehyde **10** or ketone **12** with a suitably protected amino acid derivative. Discounting subsequent protecting group manipulation, this approach would give the auxiliary-linked amino acid "cassettes" **11** and **13** ready for use in standard SPPS in a single step, and also obviate the need for the extra steps of

bromoacetylation of the solid phase peptide necessary in the "sub-monomer" approach (see **Figure 23**).

We also planned to utilise suitably labile thiol protecting groups which could be removed in the presence of glycosylation, such as the trityl (Trt), *para*-methoxybenzyl (PMB), and *ortho*-nitrobenzyl (ONB) protecting groups. The Trt group can be removed by treatment with TFA, and can therefore be removed concomitantly with side chain deprotection and cleavage of the peptide from the solid phase. The PMB and ONB groups are stable to TFA cleavage, but can be removed by oxidative cleavage by mercury acetate in the presence of acetic acid, or by photolysis, respectively. Hence these groups would be appropriate for instances where the thiol group must remain masked following cleavage from the solid phase, such as in the synthesis of peptide fragments bearing both an *N*-terminal auxiliary and a *C*-terminal thioester, as required for sequential coupling strategies. A very similar strategy utilising reductive amination and suitably labile protecting groups has since been employed by Danishefsky and co-workers.[113]

In addition to improving the efficiency and generality of the auxiliary syntheses, we were also interested in investigating some of the recent innovations in the synthesis of thioesters via the sulfonamide "safety-catch" strategy. The thioesters for the test ligations were therefore synthesised both by the traditional safety-catch protocols and also by "double-linker" methodology.[62] A number of thioester peptides were also formed using the intramolecular *N* to *S* acyl shift method previously described.[173] The utility and efficiency of these different methods were compared.

II. Results and Discussion

1. Synthesis of auxiliary peptides

We chose the *O*-linked glycoprotein GlyCAM-1 (see **Figure 24**) as a model system for the study of auxiliary-mediated ligations. In previous work in our group, the assembly of GlyCAM-1 by a sequential ligation strategy had been problematic due to the presence of a non-native cysteine residue, which was introduced into one of the glycopeptide fragments to facilitate its assembly by NCL.[118] The subsequent EPL of this fragment with a bacterially derived thioester proceeded extremely slowly, unless the non-native internal cysteine was capped by treatment with iodoacetamide to form a glutamine analogue. We envisaged that the use of auxiliary-mediated ligation would allow us to avoid this problem.

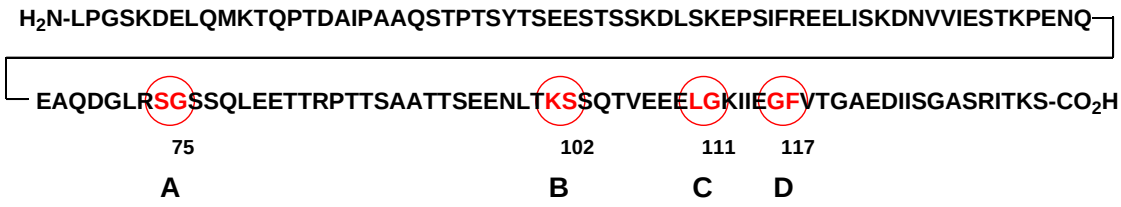


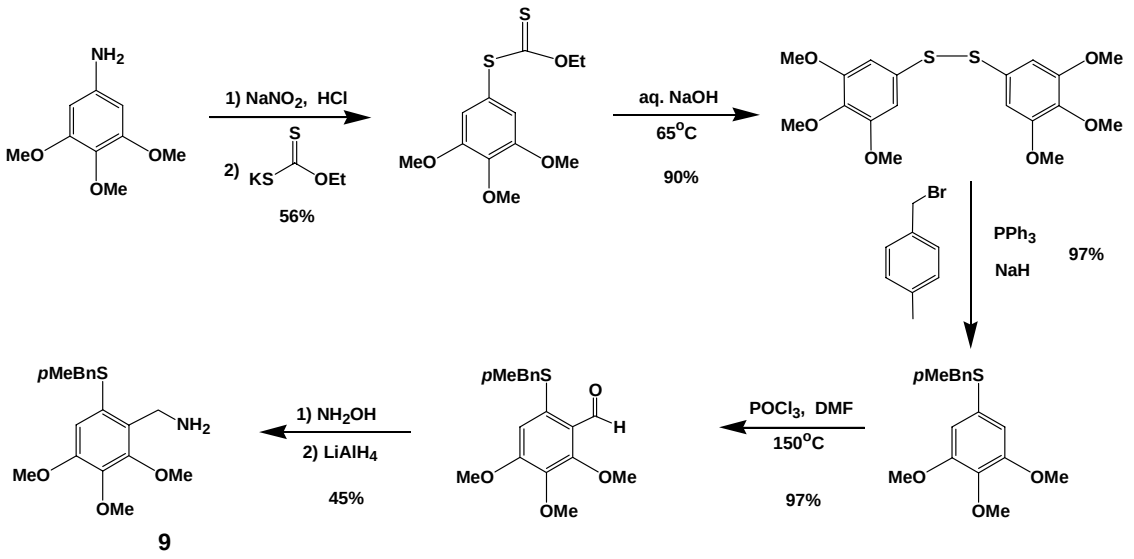
Figure 24: amino acid sequence of GlyCAM-1, with model junctions for ligation studies highlighted. **A**: model junction for Ser-Gly and Gly-Gly ligation studies; **B**: model junction for Lys-Gly ligation studies; **C**: model junction for Leu-Gly, Ala-Gly, Phe-Gly and Gly-Gly ligation studies; **D**: model junction for Gly-Ala ligation studies.

Four model ligation sites were used to study the effectiveness of auxiliary-mediated ligation for a range of ligation junctions as shown in **Figure 24**. Successful ligations at these junctions would demonstrate the feasibility of the assembly of GlyCAM-1 with minor or no changes in sequence via auxiliary-mediated ligation.

Target *N*-terminal auxiliary peptides and *C*-terminal thioester peptides were thus designed to correspond to the appropriate sequences in GlyCAM-1 and were synthesised as described in the following sections.

1.1 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary

In the original synthesis of the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary,[129] the protected thiol was constructed by displacement of *para*-methyl benzyl bromide with 3,4,5-trimethoxy-thiophenol to form the *para*-methyl benzyl thioether as shown in **Scheme 17**. This was then subjected to Vilsmeier formylation to form the aldehyde, and subsequently converted to 4,5,6-trimethoxy-2-mercaptobenzyl amine **9**, which was then introduced to the *N*-terminus of a pre-assembled peptide via the sub-monomer approach as previously described. The 3,4,5-trimethoxy-thiophenol was generated by *in situ* reduction of the corresponding disulfide, which was formed in two steps from 3,4,5-trimethoxy-aniline.

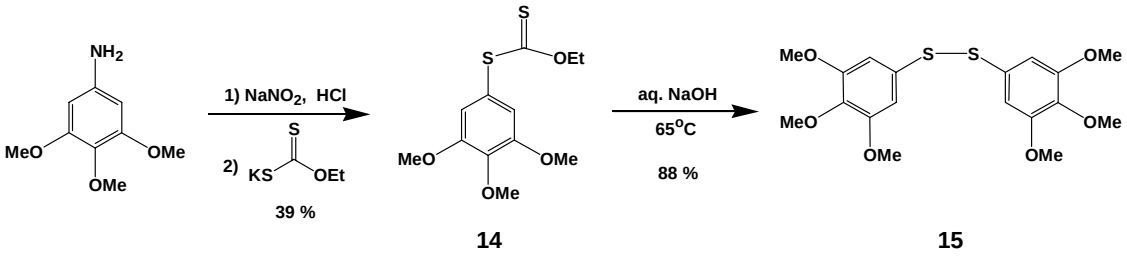


Scheme 17: original synthesis of 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary amine **9**.

To improve the generality and efficiency of the route, we planned to substitute the *para*-methyl benzyl protecting group with PMB or ONB protection by use of the appropriate benzyl halides, and also utilise a direct reductive amination with a suitably protected amino acid in the final step. If successful, this last step would offer two advantages over the original synthesis: firstly to give a probable improvement over the original low yielding amination procedure, and secondly to form an amino acid "cassette" and thus avoid "sub-monomer" conjugation to the peptide. Trityl protection for the thiol group was not investigated as this group would be unstable under the acidic conditions of the Vilsmeier reaction.

1.1.1 Synthesis of thioethers

The 3,4,5-trimethoxy-thiophenol disulfide **15** was synthesised by the procedure used in the original synthesis (see **Scheme 18**). 3,4,5-trimethoxy-aniline was subjected to diazotisation with sodium nitrite and HCl, then treated with an aqueous solution of potassium ethyl xanthate to form **14** in 39 % yield. The initial diazotisation was problematic due to the acidic aniline solution solidifying on cooling, although the reaction mixture did slowly redissolve as the $NaNO_2$ solution was added. A low yield was also reported for this procedure in the original synthesis.

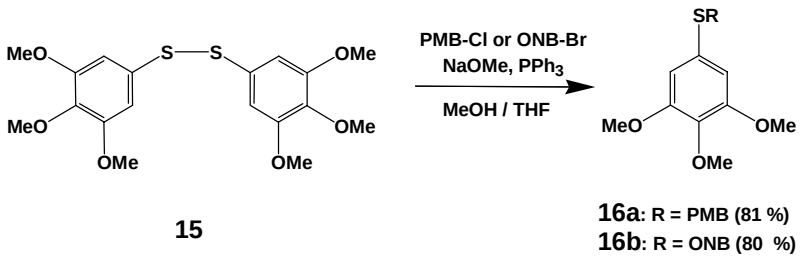


Scheme 18: synthesis of 3,4,5-trimethoxy-thiophenol disulfide **15**.

Xanthate **14** was then subjected to base hydrolysis with NaOH at 65 ^{o}C to give disulfide **15** in 88 % yield, in sufficient purity following extraction with ethyl acetate to be used directly in the next reaction. Although the expected product of hydrolysis of **14** was 3,4,5-trimethoxy-thiophenol, only the disulfide product **15** was observed, indicating rapid oxidation of the thiophenol to form the disulfide.

The PMB and ONB protecting groups were introduced in good yield by *in situ* triphenyl phosphine reduction of disulfide **15** to the corresponding 3,4,5-trimethoxy-thiophenol,

which subsequently reacted with the benzyl halide of the appropriate protecting group in the presence of sodium methoxide to form PMB thioether **16a** or ONB thioether **16b** (see **Scheme 19**).



Scheme 19: synthesis of auxiliary-thioethers **16a** and **16b**.

1.1.2 Synthesis of thioethers via Buchwald/Hartwig coupling

Although thioethers **16a** and **16b** could be successfully synthesised by the route described above, we envisaged that a shorter and more efficient synthesis could be carried out using a palladium-catalysed coupling between 3,4,5-trimethoxyphenyl bromide or triflate (**17** or **18**, see **Table 2**, p 96) and the benzyl mercaptan of the appropriate protecting group, thus forming the thioether in a single step.

This strategy would take advantage of the methodology for palladium-catalysed coupling between aryl triflates or halides and thiols or amines developed by Buchwald and co-workers[178-180] and Hartwig and co-workers[181,182] (see **Figure 25**). The 3,4,5-trimethoxyphenyl bromide and 4-methoxybenzyl mercaptan starting materials for these

reactions are commercially available, but it was necessary to synthesise 3,4,5-trimethoxyphenyl triflate **18** and 2-nitrobenzyl mercaptan **19**.

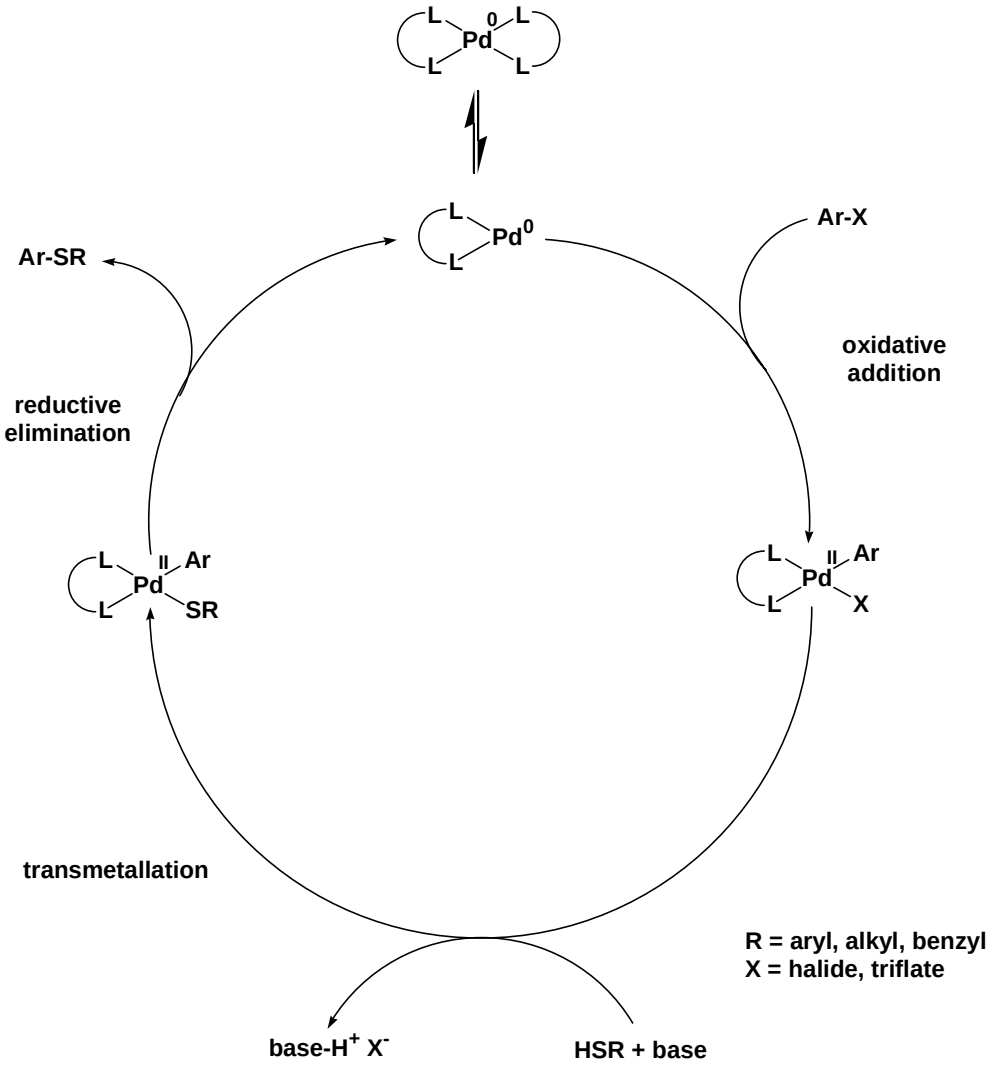
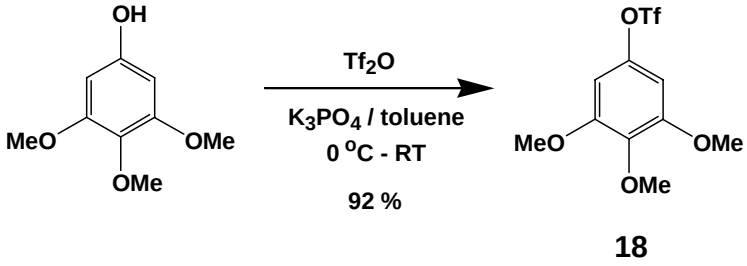


Figure 25: catalytic cycle for palladium-catalysed aryl-thiol coupling.[177]

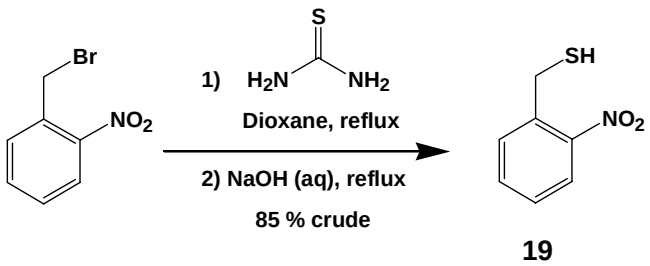
3,4,5-trimethoxyphenyl triflate **18** was synthesised in excellent yield in a straightforward manner (see **Scheme 20**). Treatment of 3,4,5-trimethoxyphenol with triflic anhydride in a 1:1 v/v solution of aqueous potassium phosphate and toluene afforded **18** following work up with no further purification required.



Scheme 20: synthesis of 3,4,5-trimethoxyphenyl triflate **18**.

Alkyl thiols can be prepared by treatment of the corresponding alkyl halide with thiourea, followed by hydrolysis.[183,184] 2-nitrobenzyl bromide was thus treated with thiourea in refluxing dioxane, followed by hydrolysis with aqueous NaOH, again under reflux (see **Scheme 21**). The reaction mixture was subsequently acidified and extracted with DCM to afford 2-nitrobenzyl mercaptan **19** in 85 % crude yield.

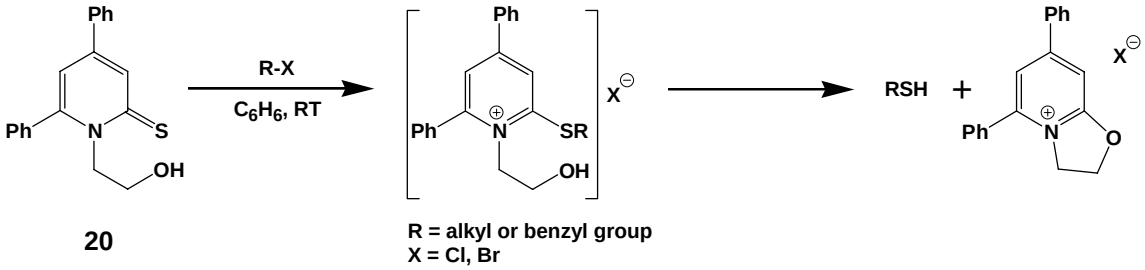
Unfortunately this product contained an impurity, which was inseparable from the product by chromatography. As palladium-catalysed couplings can be very sensitive to the presence of impurities, an alternative method for the synthesis of **19** was investigated.



Scheme 21: synthesis of **19** by treatment with thiourea followed by hydrolysis.

Treatment of alkyl or benzyl halides with 1-(2-hydroxyethyl)-4,6-diphenylpyridine-2-thione **20** has been reported by Katritzky and co-workers as a simple procedure for the formation of the corresponding thiols under mild and neutral conditions[185] (see **Scheme 22**).

Addition of alkyl or benzyl halides to a solution of **20** in dry benzene results in formation of the pyridinium cation intermediate, which undergoes intramolecular nucleophilic displacement to generate the desired thiol and the insoluble pyridinium halide salt byproduct, which can be removed by filtration. Good yields were reported for the formation of a range of alkyl and benzyl thiols by this method.



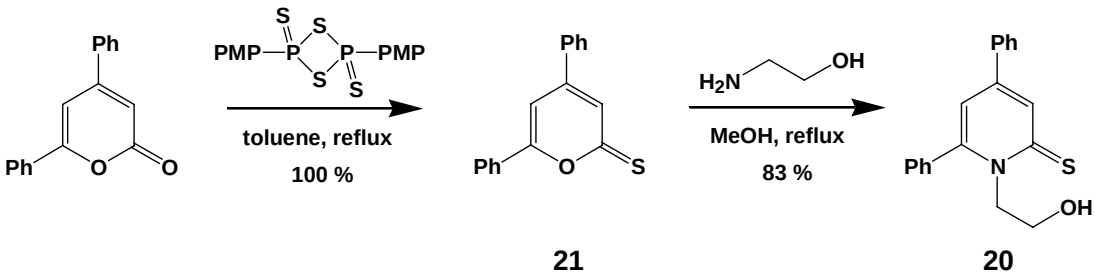
20

R = alkyl or benzyl group
X = Cl, Br

Scheme 22: thiol formation with 1-(2-hydroxyethyl)-4,6-diphenylpyridine-2-thione **20**.

Reagent **20** was synthesised in two steps as shown in **Scheme 23**. 4,6-diphenyl-2-pyrone was converted to thione **21** in quantitative yield by treatment with Lawesson's reagent in refluxing toluene.[186] **21** was then treated with ethanolamine in refluxing methanol to afford **20** in good yield. Treatment of 2-nitrobenzyl bromide with **20** at room temperature (see **Scheme 22**) afforded the desired 2-nitrobenzyl mercaptan **19** in

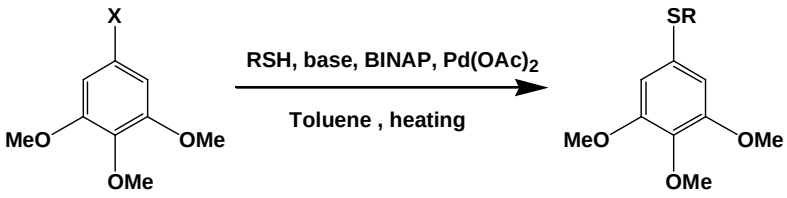
75 % yield following removal of the insoluble pyridinium halide salt by filtration and subsequent column chromatography.



Scheme 23: synthesis of 1-(2-hydroxyethyl)-4,6-diphenylpyridine-2-thione **20**.

With the required starting materials in hand, we applied published conditions[187] for palladium catalysed aryl *C – S* bond formations: heating to reflux in toluene with palladium acetate catalyst and the BINAP ligand and sodium *t*-butyloxide as base, to the coupling of triflate **18** and 4-methoxybenzyl mercaptan to form 3,4,5-trimethoxy-(4-methoxybenzyl)-thioether **16a** (**Table 2**, entry 1).

Disappointingly, negligible product formation was observed under these conditions, even after prolonged heating. When the same reaction was transferred to a microwave reactor however, heating to 120 ^{o}C at 200 W afforded **16a** in 32 % isolated yield after only 20 minutes (entry 2). Following this encouraging result, different bases which had been reported to be effective reagents for similar coupling reactions[179,187] were used in place of NaO*t*Bu (entries 3 – 4). Triethylamine proved to be the most effective base, and a gratifying yield of 74 % was achieved when the reaction with NEt_3 was conducted under an argon atmosphere (entry 5).



17: X = Br
18: X = OTf

16a: R = PMB
16b: R = ONB

Entry[a]	Substrate	Product	Base	Yield (%)
1	**18**	**16a**	NaO*t*Bu	< 5
2	**18**	**16a**	NaO*t*Bu	32
3	**18**	**16a**	10 % K_3PO_4	55
4	**18**	**16a**	NEt_3	61
5[b]	**18**	**16a**	NEt_3	74
6	**17**	**16a**	NaO*t*Bu	12
7	**17**	**16a**	NEt_3	24
8	**17**	**16a**	10 % K_3PO_4	44
9	**18**	**16b**	NEt_3	< 5
10	**17**	**16b**	10 % K_3PO_4	< 5

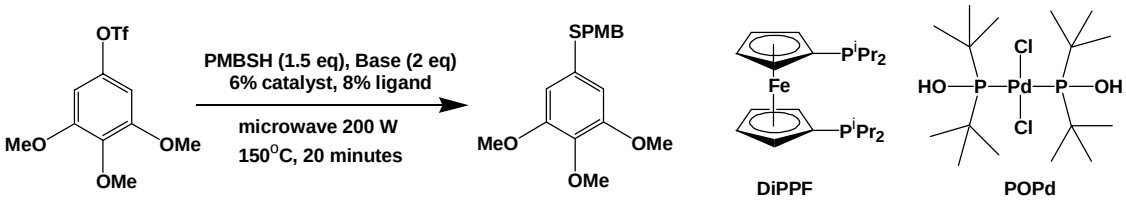
Table 2: Palladium catalysed aryl – thiol couplings. **a**) **entry 1**: 1.4 eq. thiol, 1.4 eq. NaO*t*Bu, 11 mol % $Pd(OAc)_2$, 11 mol % BINAP, 3 eq. LiCl, 80 ^oC, 24 hours; **entries 2-10**: 1.4 eq. thiol, 1.4 eq. base, 11 mol % $Pd(OAc)_2$, 11 mol % BINAP, 120 ^oC microwave heating, 200 W, 20 minutes; **b**) reaction performed under Ar atmosphere. BINAP: 1,1-binaphthalene-2,2-diylbis-(diphenylphosphine)

Commercially available 3,4,5-trimethoxyphenyl bromide **17** proved to be a poorer substrate for the coupling reaction than triflate **18** (entries 6 – 8). The best results for coupling of **17** were obtained with aqueous potassium phosphate as the base (entry 8). Unfortunately, the best coupling conditions for both **17** and **18** were unsuccessful for

their coupling with 2-nitrobenzyl mercaptan **19** to form 3,4,5-trimethoxy-(2-nitrobenzyl)-thioether **16b** (entries 9 – 10). Electron-deficient thiols such as **19** have previously been reported to be poor substrates for these coupling reactions.[182]

Further optimisation studies for the formation of thioether **16a** from triflate **18**, using a range of different solvents, bases and catalysts, were performed using an automated parallel microwave reactor and HPLC analysis. In addition to those reagents previously studied, the 1,1-bis-(diisopropylphosphino)-ferrocene (DiPPF)[180] and PPh_3 ligands, the catalyst $[(t\text{Bu})_2\text{P(OH)}]_2\text{PdCl}_2$ (POPd)[188] and Cs_2CO_3 as base were also investigated The trial reactions were carried out under the conditions shown and the results are shown in **Table 3**. The most significant results are highlighted.

An increase in reaction temperature to 150 ^{o}C resulted in a small increase in yield over the previous best conditions. More dramatic results were observed on changing the solvent, with DMF and dioxane particularly effective. DiPPF and triethylamine was found to be the most effective ligand-base combination, affording thioether **16a** in near-quantitative yield with dioxane as solvent. Triethylamine was again the most effective base for each of the three catalyst systems used. The POPd catalyst gave poor yields of **16a** in the presence of NEt_3, and no observable product with the other bases. Triphenylphosphine was completely ineffective as a ligand in all trials (results not tabulated). These two results confirm the previously reported importance of aromatic ligands for these coupling reactions[182] and in particular the requirement for bidentate chelating ligands.[178,182]



18 **16a**

Catalyst	Ligand	Base	Solvent				
			toluene	DMF	dioxane	MeCN	H_2O
			Yield (%)				
$Pd(OAc)_2$	BINAP	NEt_3	76	85	78	36	0
$Pd(OAc)_2$	BINAP	NaO*t*Bu	28	48	58	53	39
$Pd(OAc)_2$	BINAP	K_3PO_4	3	5	48	36	33
$Pd(OAc)_2$	BINAP	Cs_2CO_3	0	4	67	0	0
$Pd(OAc)_2$	DiPPF	NEt_3	28	90	97	92	28
$Pd(OAc)_2$	DiPPF	NaO*t*Bu	16	0	59	0	47
$Pd(OAc)_2$	DiPPF	K_3PO_4	10	15	79	18	50
$Pd(OAc)_2$	DiPPF	Cs_2CO_3	7	25	74	13	51
POPd	-	NEt_3	4	37	30	33	17
POPd	-	NaO*t*Bu	0	0	0	0	0
POPd	-	K_3PO_4	0	0	0	0	0
POPd	-	Cs_2CO_3	0	0	0	0	0

Table 3: HPLC results for coupling optimisation studies.

Use of the best conditions from these trials (DiPPF with NEt_3 in dioxane) on a preparative scale afforded **16a** in a yield of 97 %, identical to that observed by HPLC analysis. Very similar results have subsequently been reported for these couplings under microwave conditions.[189] Formation of the ONB thioether **16b** with the optimised conditions for the formation of **16a** was not attempted to due to problems encountered with the ONB protecting group at a later stage (see **Section 1.1.4**).

1.1.3 Synthesis of auxiliary-amino acid cassettes

Following formation of PMB and ONB thioethers **16a** and **16b**, the next step was preparation of the corresponding auxiliary aldehydes **10a** and **10b** as precursors for the key direct reductive amination. **16a** and **16b** were thus subjected to Vilsmeier formylation. We were initially concerned that the PMB protecting group might be unstable to the HCl byproduct generated by Vilsmeier formylation (see **Figure 26**) under the high temperatures (150 ^{o}C) used in the original synthesis.[129]

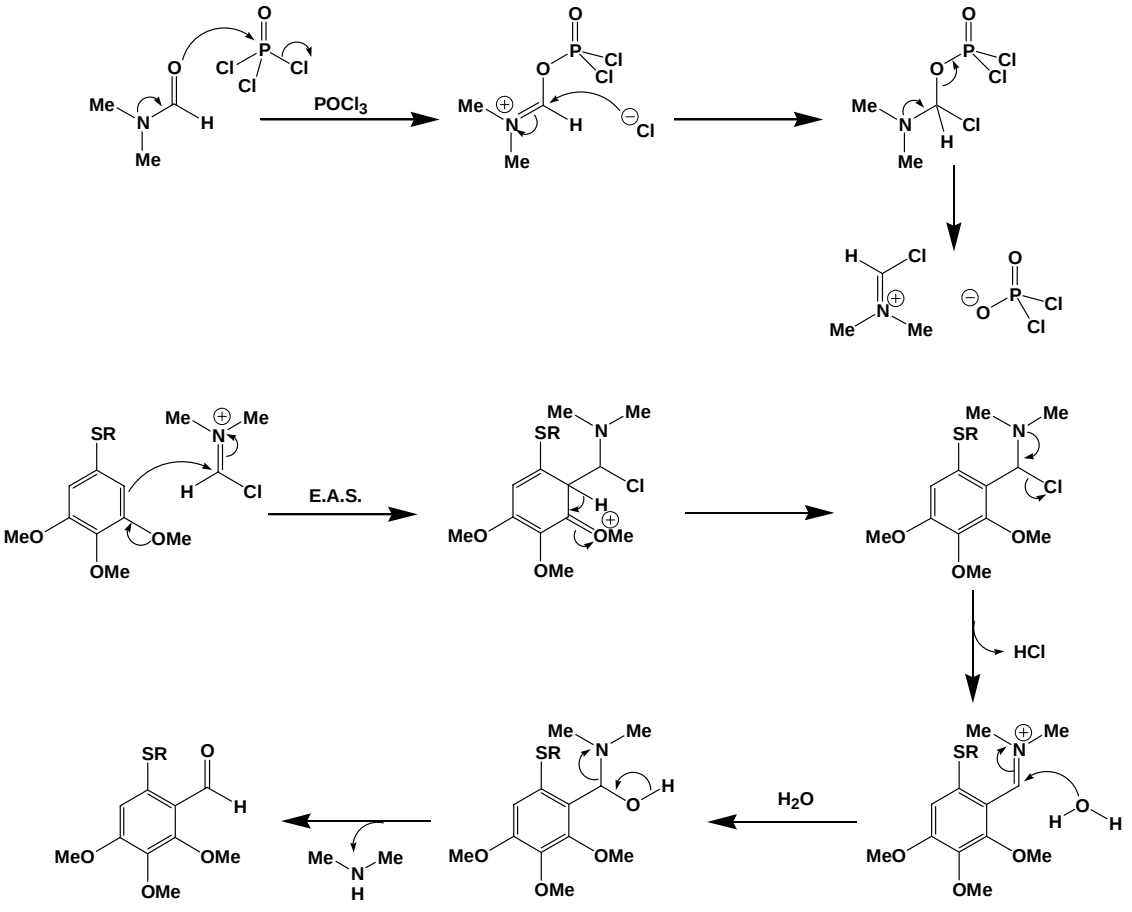
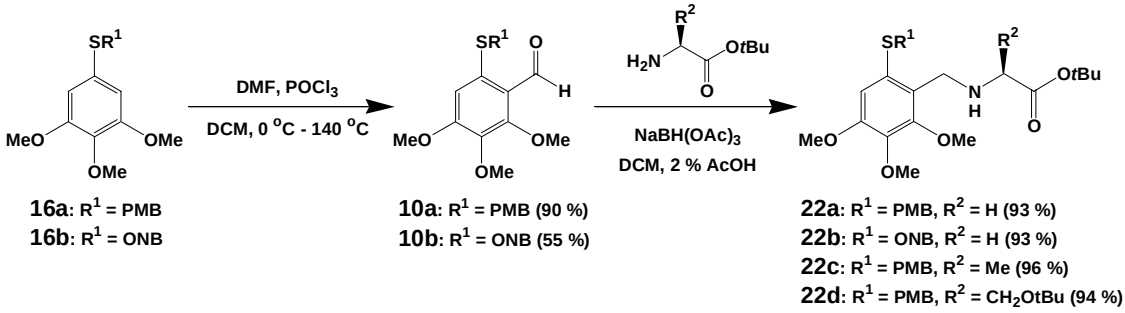


Figure 26: mechanism of Vilsmeier formylation.

The reaction was therefore initially conducted at a lower temperature (90 ^{o}C) for a longer duration of 16 hours with PMB thioether **16a**. Unfortunately this procedure afforded 4,5,6-trimethoxy-2-(4-methoxybenzylthio)-benzaldehyde **10a** in only 49 % yield. Heating at 150 ^{o}C as in the original synthesis afforded **10a** in an excellent yield of 90 %, although longer reaction times for both the formation of the intermediate iminium cation and the hydrolysis step were necessary. Care was required to avoid evaporation of the DCM solvent to dryness under the high reaction temperatures, as best results were obtained when the minimum volume of solvent was used to ensure a highly concentrated reaction. Saturated sodium bicarbonate solution was used in the hydrolysis step to minimise the chance of product degradation due to the HCl byproduct. ONB thioether **16b** was a less good substrate for this reaction due to its poorer solubility in DCM. The high concentration necessary for an effective reaction was therefore more difficult to achieve. As a result of these difficulties, ONB aldehyde **10b** was formed in a lower yield of 55 % (see **Scheme 24**).



16a: R^1 = PMB
16b: R^1 = ONB

10a: R^1 = PMB (90 %)
10b: R^1 = ONB (55 %)

22a: R^1 = PMB, R^2 = H (93 %)
22b: R^1 = ONB, R^2 = H (93 %)
22c: R^1 = PMB, R^2 = Me (96 %)
22d: R^1 = PMB, R^2 = CH_2OtBu (94 %)

Scheme 24: synthesis of *t*-butyl protected auxiliary-amino acids **22a – 22d**.

Auxiliary aldehydes **10a** and **10b** were then subjected to reductive amination with the *t*-butyl esters of glycine, alanine and serine in the presence of sodium

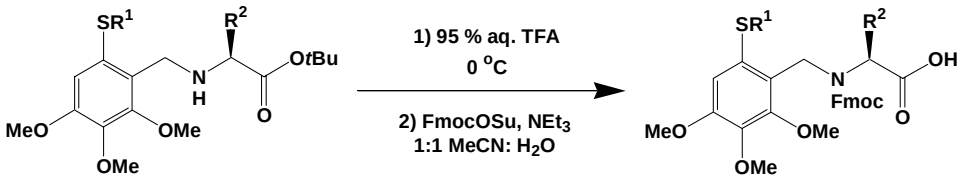
triacetoxyborohydride in DCM with 2 % acetic acid added as a catalyst (see **Scheme 24**). The *t*-butyl protected acids were used to aid subsequent purification.

All of these reactions were practically very straightforward and proceeded rapidly to give protected auxiliary-amino acids **22a** – **22d** in excellent yields after simple purification by column chromatography. This was especially gratifying in comparison with the poor yield for the formation of the 4,5,6-trimethoxy-2-mercaptobenzyl amine **9** via a prolonged two-step procedure at this stage in the original synthesis.

The final step in the synthesis was protecting group manipulation to remove the *t*-butyl ester protection (and the *t*-butyl ether side chain protecting group for serine) and install the Fmoc protecting group to give the auxiliary-amino acid "cassettes" ready for conjugation to a peptide by standard SPPS. We planned to accomplish this by removal of the *t*-butyl protecting groups with aqueous TFA, followed by treatment with Fmoc-succinimide in the presence of triethylamine in a 1:1 v/v mixture of acetonitrile and water to install the Fmoc protecting group.

This apparently straightforward protecting group manipulation proved problematic, however. Although TFA deprotection went to completion in each case as monitored by TLC, the Fmoc protection step proceeded very sluggishly, with prolonged reaction times and a second addition of reagents necessary. A column chromatography step was also necessary to isolate the Fmoc-protected products in sufficient purity. The Fmoc-protected auxiliary-glycine cassettes **11a** and **11b** were formed in 67 % and 71 % yield respectively (see **Scheme 25**), but Fmoc protection of the more sterically bulky alanine

and serine derivatives derived from **22c** and **22d** was extremely difficult. The Fmoc-protected alanine derivative was isolated in only 8 % yield after overnight reaction, and the Fmoc-protected serine derivative could not be isolated.



22a: R^1 = PMB, R^2 = H
22b: R^1 = ONB, R^2 = H
22c: R^1 = PMB, R^2 = Me
22d: R^1 = PMB, R^2 = CH_2OtBu

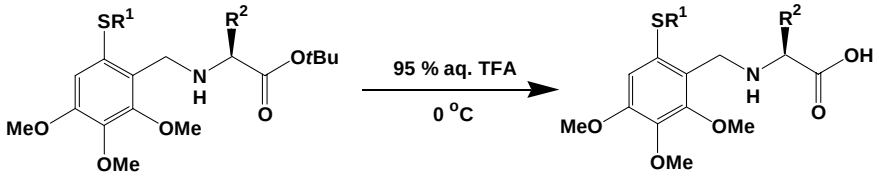
11a: R^1 = PMB, R^2 = H (67 %)
11b: R^1 = ONB, R^2 = H (71 %)

Scheme 25: synthesis of Fmoc-protected auxiliary-glycine cassettes **11a** and **11b**.

The purpose of the Fmoc protecting group was to ensure that byproducts arising from the acylation of auxiliary-amino acids with the activated esters of the same species were not formed during conjugation to the solid phase peptide. However, as the steric bulk around the secondary amines of alanine and serine appeared to inhibit acylation of these auxiliary-amino acids with Fmoc-succinimide, it seemed likely that acylation with the activated ester of a second molecule of auxiliary-amino acid would suffer the same inhibitory effect, and thus Fmoc protection might not be required at all for these compounds.

N-4,5,6 - trimethoxy- 2-(4-methoxybenzylthio)benzyl alanine **11c** and *N*-4,5,6 - trimethoxy- 2-(4-methoxybenzylthio)benzyl serine **11d** were therefore isolated by column chromatography following *t*-butyl deprotection in 76 % and 71 % yields respectively (see **Scheme 26**). These somewhat low yields were mainly due to the

practical difficulties associated with chromatographic isolation of the polar free acid compounds, as TLC analysis indicated quantitative or near-quantitative reaction.



22c: R^1 = PMB, R^2 = Me
22d: R^1 = PMB, R^2 = CH_2OtBu

11c: R^1 = PMB, R^2 = Me (76 %)
11d: R^1 = PMB, R^2 = CH_2OH (71 %)

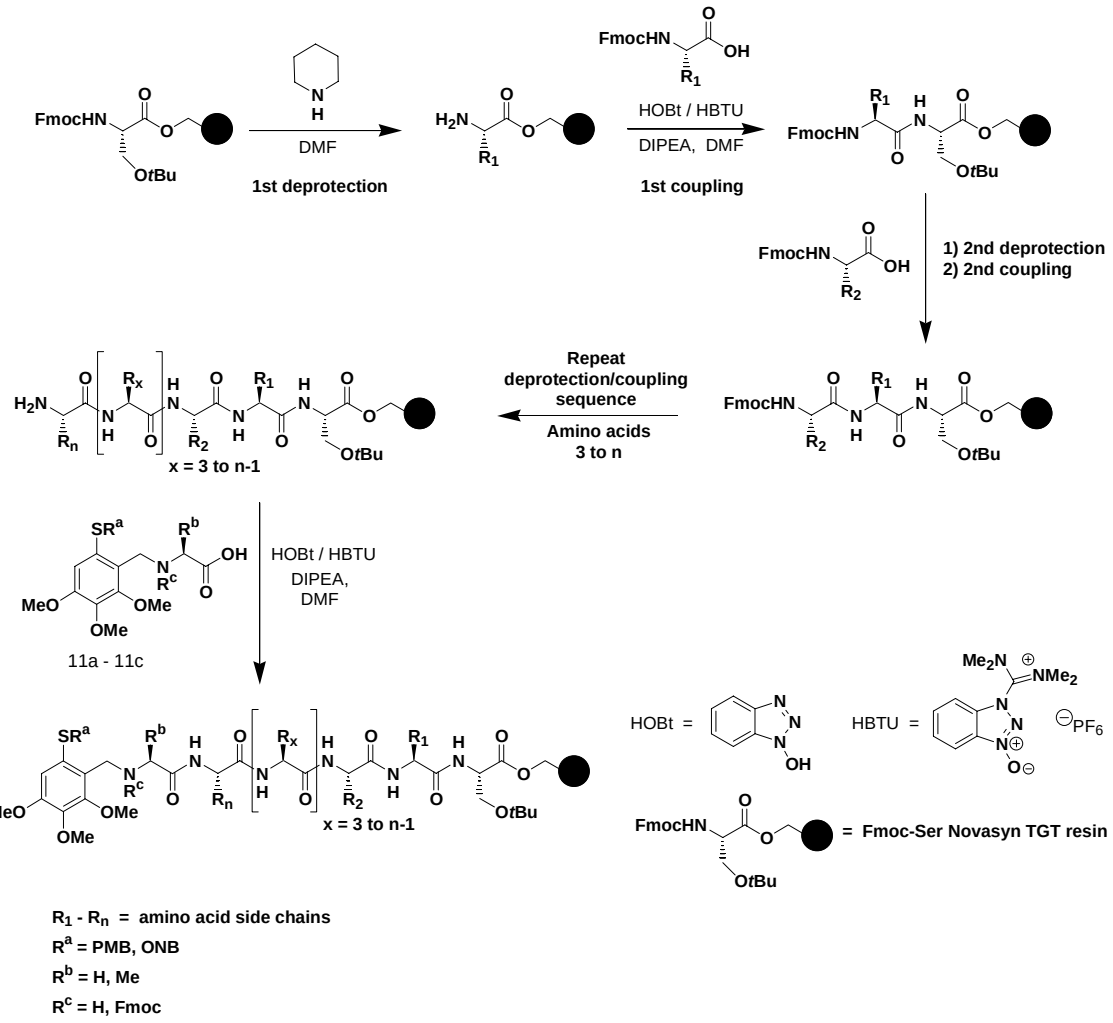
Scheme 26: synthesis of auxiliary-alanine cassette **11c** and auxiliary-serine cassette **11d**.

PMB protected auxiliary-amino acid cassettes **11a**, **11c** and **11d** ready for use in SPPS were thus synthesised in four steps in good overall yields (54 % - 64 %) utilising a highly efficient novel palladium catalysed aryl – thiol coupling reaction and direct reductive amination.

This route therefore represents a considerable improvement over the original synthesis, being shorter and higher yielding and compatible with sensitive peptide modifications due to the use of the PMB protecting group, and it also avoids the need for "sub-monomer" peptide conjugation and the associated extra steps. ONB-protected auxiliary-amino acid cassettes such as **11b** could also be formed by a longer route. Optimisation of the synthesis of ONB-protected auxiliaries was not carried out due to problems subsequently encountered during attempted removal of the ONB protecting group (see section **1.1.4**).

1.1.4 Synthesis of auxiliary-peptides

Peptides corresponding to fragments of the glycoprotein GlyCAM-1 were synthesised on NovaSyn®TGT resin preloaded with serine with a loading of 0.22 mmol/g (see **Scheme 27**).



Scheme 27: Fmoc SPPS of auxiliary peptides.

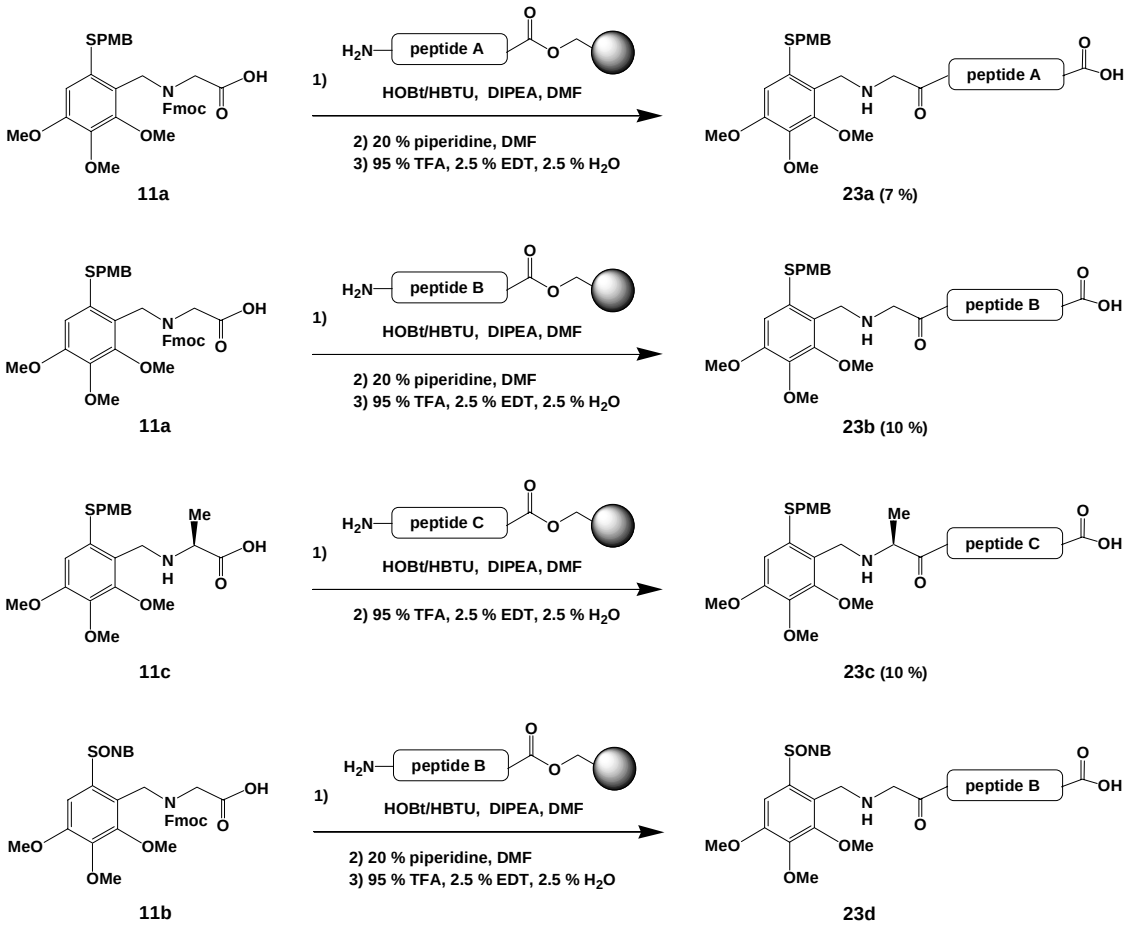
Synthesis was conducted on a 0.05 mmol scale by manual Fmoc SPPS, using 10 equivalents of each Fmoc amino acid per coupling and HBTU/HOBt coupling reagents

in the presence of Hünigs base. Fmoc protecting group removal was carried out with 20 % v/v piperidine in DMF. Synthesis was monitored by the Kaiser ninhydrin test[190] and LC-MS analysis of samples of the peptide.

Auxiliary-amino acid cassettes **11a** – **11c** were coupled to peptides **A** - **C**, corresponding to fragments of GlyCAM-1, using five equivalents of amino acid and coupling reagents (see **Scheme 28**). These couplings went to completion as monitored by LC-MS. This result was especially significant in the case of auxiliary-alanine cassette **11c** as the reported formation of an *N*-terminal alanine peptide bearing the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary by the sub-monomer approach occurred in only 50 % yield.[129] No evidence of coupling of auxiliary dipeptides or other byproducts resulting from acylation of the secondary amine of **11c** was detected, confirming that Fmoc protection was not necessary for the auxiliary-alanine cassette.

Following coupling of auxiliary-glycine cassettes **11a** and **11b** to peptides **A** and **B**, the Fmoc protecting groups were removed from the resulting solid phase peptides by treatment with 20 % v/v piperidine in DMF as before (see **Scheme 28**). The resulting peptides were then cleaved from the resin with concomitant side chain deprotection by treatment with 95 % v/v TFA, with 2.5 % v/v ethanedithiol (EDT) as a scavenger reagent. Following coupling of auxiliary-alanine cassette **11c** to peptide **C**, the resulting peptide could be cleaved directly from the solid phase, as the Fmoc deprotection step was not necessary in this case.

Each of the crude peptides was then purified by semi-preparative reverse-phase HPLC to afford peptides **23a** - **23c** bearing the PMB-protected auxiliary, and peptide **23d** bearing the ONB-protected auxiliary, in 7 % - 10 % overall yield (see **Scheme 28**).



Scheme 28: synthesis of PMB-protected auxiliary peptides **23a** - **23c** and ONB-protected auxiliary peptide **23d**.

The semi-preparative HPLC trace for the purification of PMB-protected auxiliary-glycine peptide **23b** and the electrospray (ESI) mass spectrum for the isolated product

are shown in **Figure 27**. The fragment ion peak due to cleavage of the auxiliary at the benzylic position appears to be characteristic of the auxiliary species and was generally observed for auxiliary-amino acids and auxiliary-peptides.

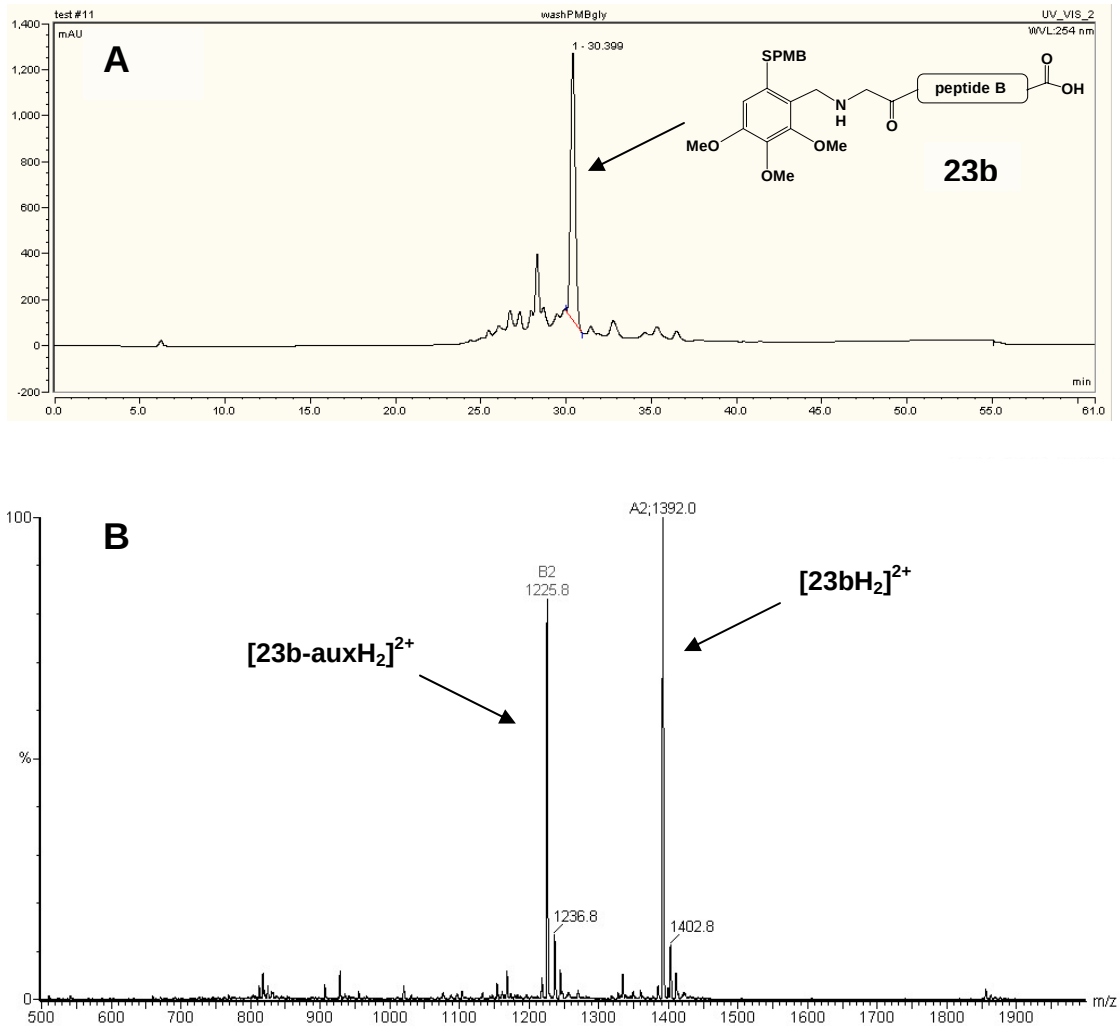
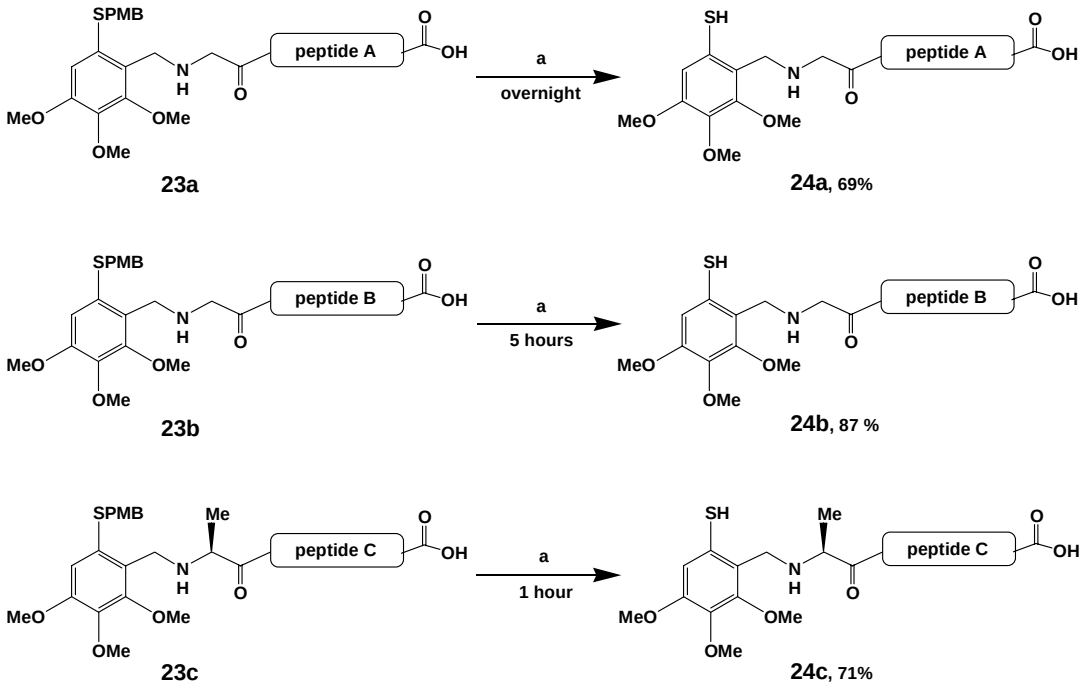


Figure 27: isolation of auxiliary-glycine peptide **23b**. **A**: 254 nm UV trace of HPLC purification; **B**: ESI positive ion mass spectrum of purified **23b** (calculated mass = 2782.2 Da, observed: 1392.0 Da $[MH_2]^{2+}$, 1225.8 Da $[M\text{-}AuxH_2]^{2+}$) showing characteristic fragmentation of auxiliary at the benzylic position. peptide B = KIIEGFVTGAEDIISGASRITKS.

The PMB protecting group was removed from peptides **23a** – **23c** by treatment with excess mercury(II)acetate in 10 % v/v aqueous acetic acid. The reaction mixture was quenched by the addtion of solid dithiothreitol (DTT) to a final concentration of 5 % w/v and shaking for one hour, after which the thick white precipitate was removed and the resulting solution was purified by semi-preparative HPLC to afford peptides **24a** - **24c** bearing the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary.



peptide A: GlyCAM-1 101-132 = SQTVEEELGKIIEGFVTGAEDIISGASRITKS

peptide B: GlyCAM-1 110-132 = KIIEGFVTGAEDIISGASRITKS

peptide C: GlyCAM-1 116-132 = VTGAEDIISGASRITKS

Scheme 29: Removal of PMB protecting group: **a**) 10 % v/v AcOH, $Hg(OAc)_2$, followed by 5 % w/v DTT.

A direct relationship was observed between the size of the peptide and the time taken for completion of the reaction. Deprotection of the shorter peptide **23c** was complete

when quenched after one hour, whereas overnight reaction was required for the deprotection of the larger peptide **23a** (see **Scheme 29**). The HPLC trace for the purification of auxiliary-peptide **4b** and the ESI mass spectrum for the isolated product are shown in **Figure 28**.

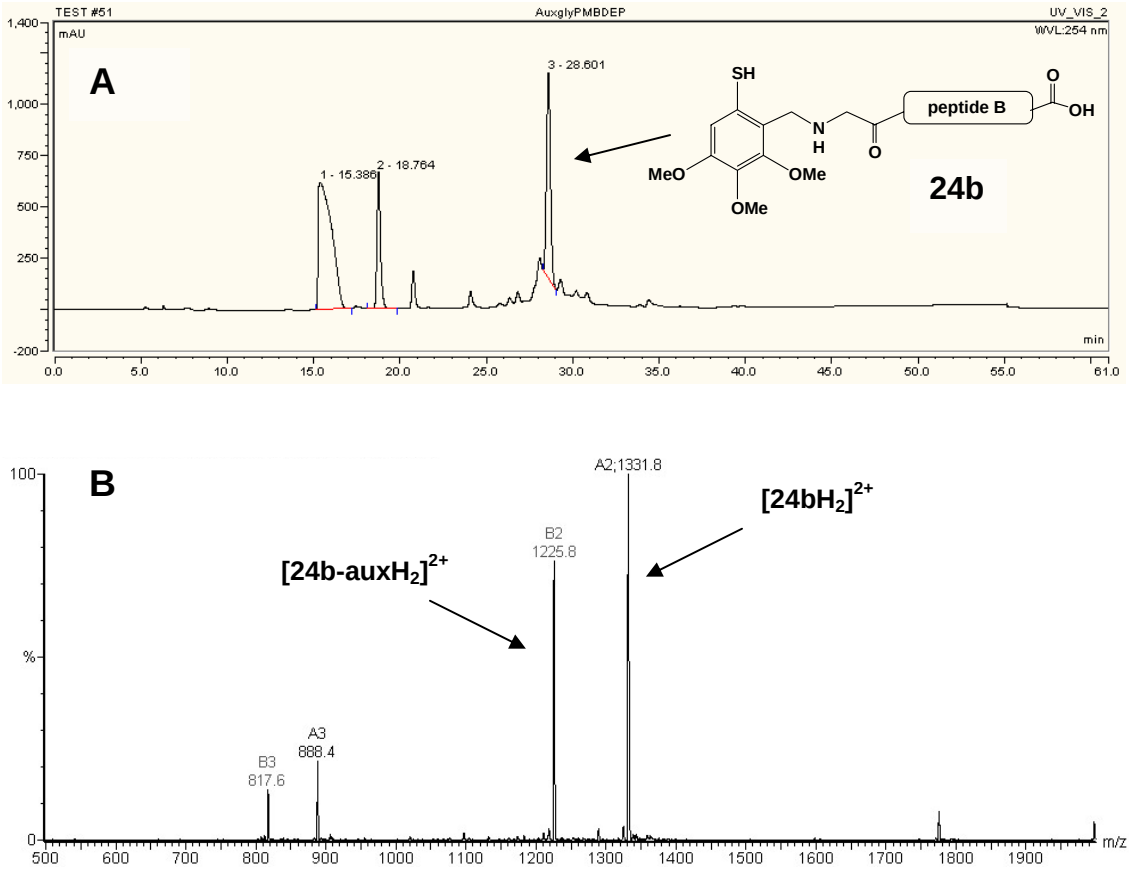
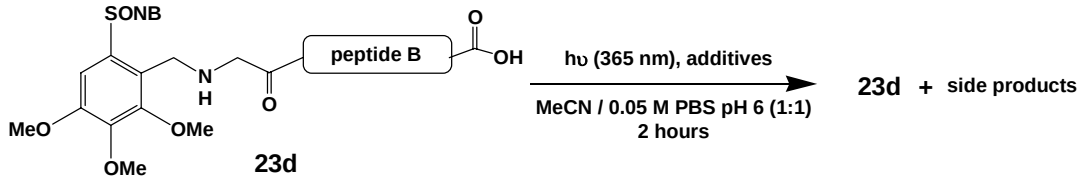


Figure 28: isolation of auxiliary-glycine peptide **24b**. **A**: 254 nm UV trace of HPLC purification; **B**: ESI positive ion mass spectrum of purified **24b** (calculated mass = 2662.02 Da, observed: 1331.8 Da $[MH_2]^{2+}$, 1225.8 Da $[M\text{-}AuxH_2]^{2+}$, 888.4 Da $[MH_3]^{3+}$, 817.6 Da $[M\text{-}AuxH_3]^{3+}$). peptide B = KIIEGFVTGAEDIISGASRITKS.

Deprotection of ONB-protected auxiliary-glycine peptide **23d** was attempted by UV photolysis under published conditions for the removal of this protecting group from cysteine.[192] A degassed solution of the peptide in 1:1 v/v acetonitrile/PBS buffer (pH 6.0), containing L-(–)-ascorbic acid as an antioxidant and semicarbazide hydrochloride as a carbonyl scavenger, was subjected to photolysis at 365 nm for two hours (see **Scheme 30**). Unfortunately, no evidence of the deprotected product was observed by mass spectrometry. Instead only the ONB protected starting material **23d** was observed**,** along with multiple similar signals assumed to be due to peptide degradation. Resubjection for a further two hours merely resulted in further degradation of **23d**.



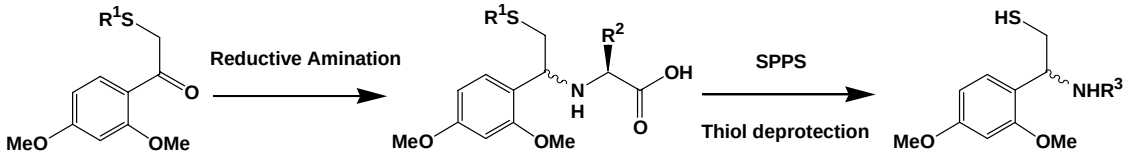
Scheme 30: attempted photolytic removal of ONB protecting group from auxiliary-glycine peptide **23d**.

In summary, although the failure of the photolytic ONB deprotection represented a setback, 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary-peptides could be efficiently assembled using PMB-protected building blocks, which could be rapidly synthesised in high yield via: 1) palladium catalysed aryl – thiol coupling; 2) Vilsmeier formylation; 3) direct reductive amination with amino acid *t*-butyl ester; 4) protecting group manipulation. PMB-protected auxiliary-glycine and auxiliary-alanine building blocks were efficiently incorporated into peptides using standard SPPS. Following cleavage

from the solid phase, removal of the PMB protecting group with mercury acetate in acetic acid afforded the auxiliary-peptides **24a** – **24c** ready for use in cysteine-free peptide ligation. Results for the ligation and auxiliary removal studies conducted with these substrates are presented in section **3**.

1.2 *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary

Peptides bearing the *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** were constructed according to the strategy shown in **Figure 29**. The ketones **12a** and **12b**, containing the protected thiol group, were formed by displacement of the commercially available 2-bromo-2,4-dimethoxyacetophenone with the appropriate thiol.



12a: R^1= PMB
12b: R^1= Trt

13a: R^1= PMB, R^2 = H
13b: R^1= Trt, R^2 = H
13c: R^1= Trt, R^2 = Me

6a-6d: R^3 = peptide

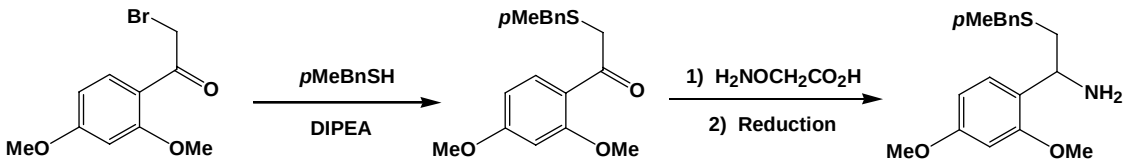
Figure 29: synthetic strategy for *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary peptides

Direct reductive amination with the appropriate amino acid afforded the auxiliary-amino acid cassettes **13a** - **13c**, which were incorporated into peptides by standard SPPS. It was possible to use unprotected amino acids for the reductive amination and hence

avoid a further deprotection step. Alternatively, the ketone could be converted to the primary amine by treatment with ammonium bicarbonate to form the *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl amine for conjugation to a peptide via the submonomer approach. Following SPPS, the thiol protecting group was removed to afford the desired peptides **6a** – **6d**.

1.2.1 Synthesis of auxiliary-amino acid cassettes and amines

The original route for the synthesis of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)-peptides[132] was reported in very little detail (see **Scheme 31**). It involved displacement of 2-bromo-2,4-dimethoxyacetophenone with *para*-methylbenzyl mercaptan to form 2-keto-2-(2,4-dimethoxyphenyl) (*S*-paramethylbenzyl) ethanethiol, followed by a two step amination-reduction procedure to afford the auxiliary benzylamine 2-amino-2-(2,4-dimethoxyphenyl) (*S*-paramethylbenzyl) ethanethiol, which was conjugated to a peptide via the submonomer approach.



Scheme 31: original synthesis of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl amine.

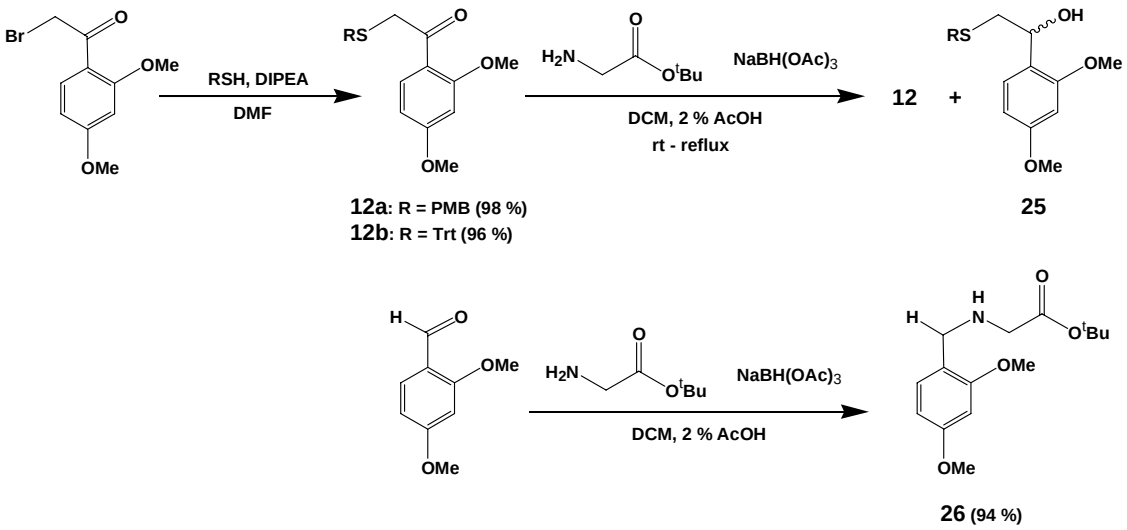
The more recently published route to *N*α-(1-(2-nitrophenyl)-2-mercaptoethyl)-peptides[135] involves a multistep synthesis of the trityl-protected auxiliary-benzylamine,

which is again used for submonomer conjugation to a peptide. We envisaged that a more efficient synthesis would involve introduction of the protected thiol by displacement of 2-bromo-2,4-dimethoxyacetophenone as in the original synthesis, with the replacement of the *para*-methylbenzyl group with PMB or trityl protection to allow the use of deprotection conditions compatible with glycosylated peptides. A single-step direct reductive amination with the appropriate amino acid would then afford the auxiliary-amino acid cassette **13** for incorporation into a peptide by standard SPPS, as for the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4**.

PMB- and trityl-protected ketones 2-keto-2-(2,4-dimethoxyphenyl) (*S*-paramethoxybenzyl) ethanethiol **12a** and 2-keto-2-(2,4-dimethoxyphenyl) (*S*-trityl) ethanethiol **12b** were formed in excellent yield by straightforward displacement of commercially available 2-bromo-2,4-dimethoxyacetophenone with the appropriate thiol reagent (see **Scheme 25**). The ONB protecting group was not investigated due to the problems experienced with its removal (see section **1.1.4**).

Disappointingly, the conditions which had afforded the 4,5,6-trimethoxy-2-mercaptobenzyl-amino acids **22a** – **22d** in high yield were unsuccessful for the less reactive acetophenones **12a** and **12b** (see **Scheme 32**). No reaction was observed on stirring for four hours at room temperature. Increasing the reaction temperature afforded a mixture of the starting ketone and the 2-(4-methoxybenzyl)-2,4-dimethoxyacetophenyl alcohol **25** arising from ketone reduction by sodium triacetoxyborohydride. The large difference in reactivity between benzaldehydes and the acetophenones was illustrated by the successful reductive amination of the related

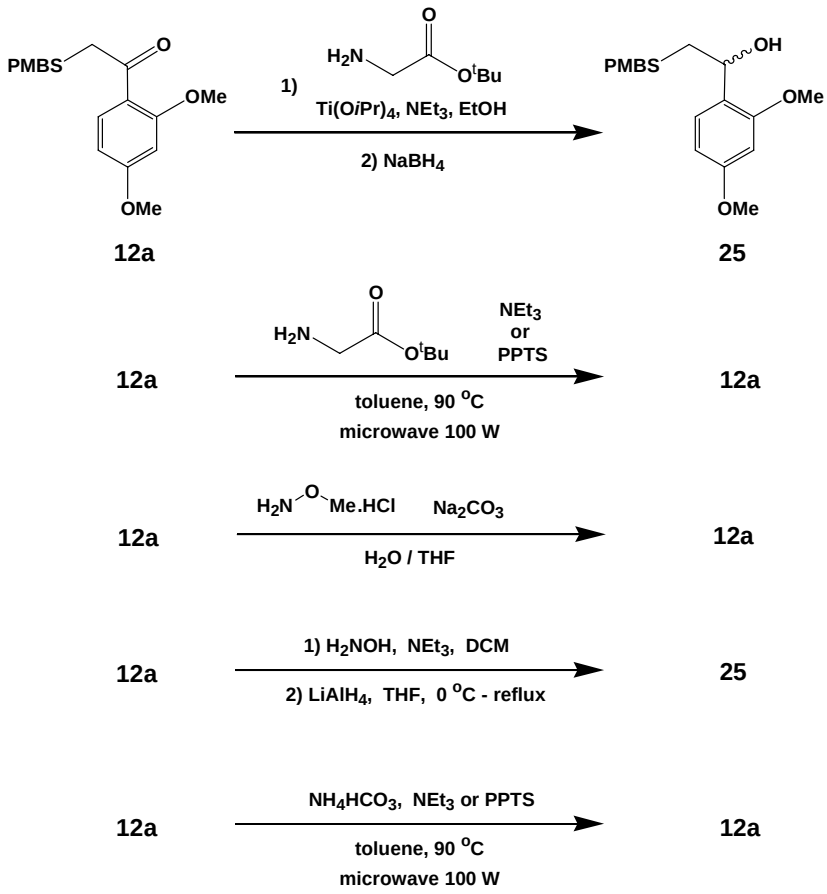
2,4-dimethoxybenzaldehyde under identical conditions to afford *N*-2,4-dimethoxybenzylglycine *t*-butyl ester **26** in 94 % yield.



Scheme 32: formation of ketones **12a** and **12b** and attempted reductive amination.

A range of different reductive amination conditions were attempted for the formation of the auxiliary-amino acid cassette (see **Scheme 33**). A two-step procedure involving Lewis-acid catalysed amination followed by reduction with the stronger reducing agent sodium borohydride resulted in formation of alcohol **25** only. The use of higher temperatures and microwave heating in the presence of either an acid or base catalyst was also unsuccessful, as only the starting ketone **12a** was observed by TLC and MS analysis in both cases. Conversion of the ketone to the corresponding oxime with methoxylamine as a precursor of the auxiliary-benzylamine, as in the original synthesis, was also unsuccessful, as were attempts to form the primary benzylamine via two-step amination-reduction and microwave-assisted amination, which resulted in formation of

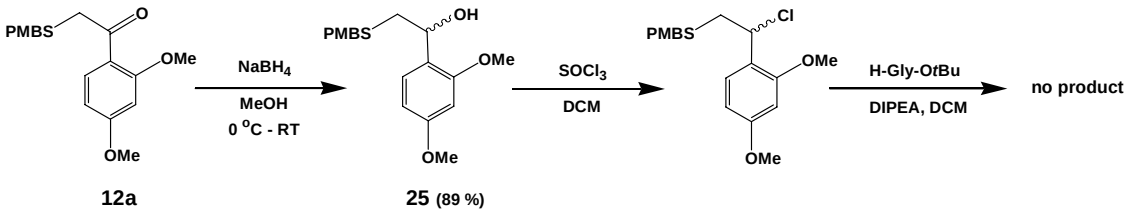
alchol **25**, and no reaction, respectively. Similar difficulties with the reductive amination of methoxyphenyl ketones have been reported by Clive and co-workers[193]



Scheme 33: unsuccessful reductive aminations. PPTS: pyridinium *para*-toluenesulfonate.

The difficulties encountered with the reductive amination of the **12a** prompted us to consider alternative methods for the introduction of the amino acid. We therefore examined the possibility of formation of the auxiliary-amino acid cassette via displacement of the benzyl halide or sulfonate. **12a** was first reduced to racemic alcohol **25** in high yield by treatment with sodium borohydride. This was then converted into 2-

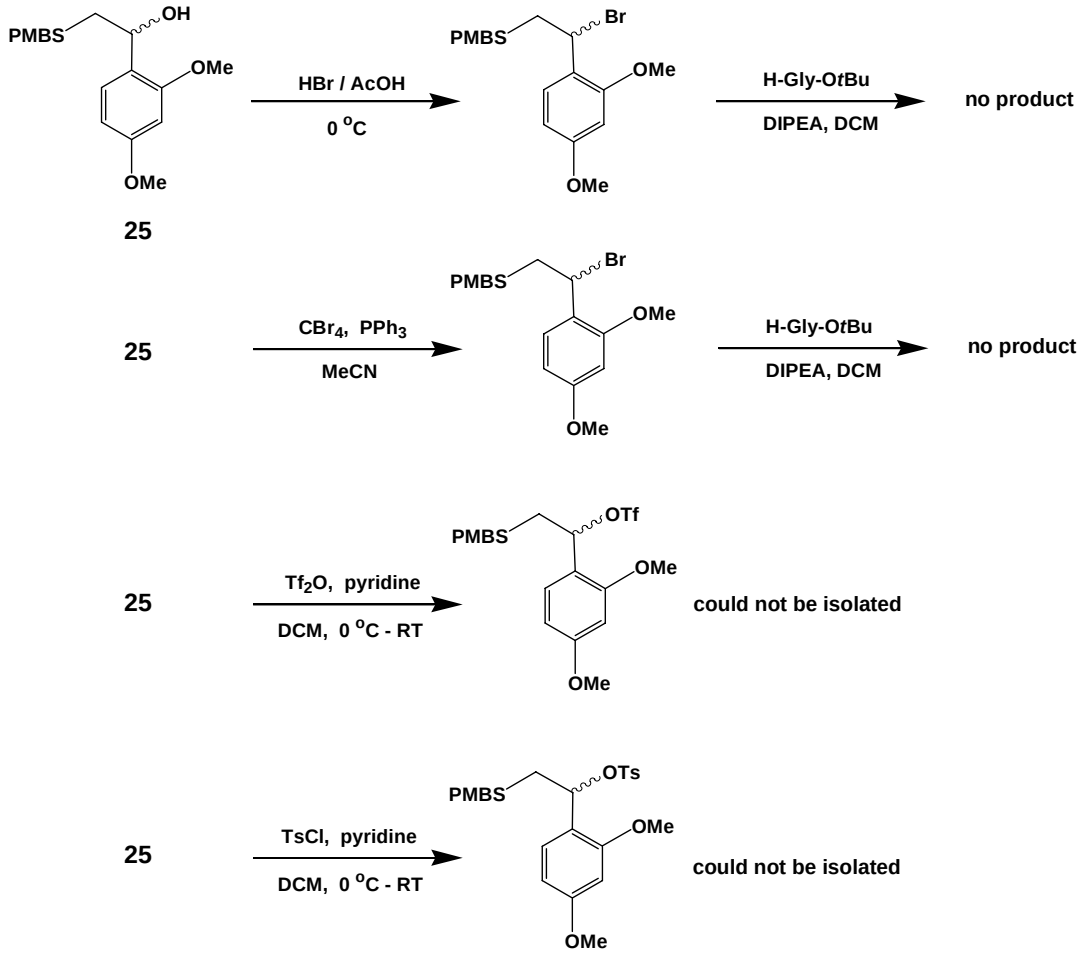
chloro-2-(2,4-dimethoxyphenyl) (*S*-paramethoxybenzyl) ethanethiol by reaction with thionyl chloride in DCM (see **Scheme 34**). As the product was unstable to chromatography, the crude product obtained following work up was treated directly with glycine *t*-butyl ester in the presence of Hünigs base. No product could be isolated from this reaction, however. It is likely that decomposition of the unstable benzyl chloride had occurred before the displacement could take place.



Scheme 34: formation of 2-chloro-2-(2,4-dimethoxyphenyl) (*S*-paramethoxybenzyl) ethanethiol and attempted displacement.

Unfortunately 2-bromo-2-(2,4-dimethoxyphenyl) (*S*-paramethoxybenzyl) ethanethiol, prepared by treatment of **25** with either hydrobromic acid in acetic acid, or with carbon tertrabromide in the presence of triphenylphosphine, as shown in **Scheme 35**, was similarly unstable. Although alcohol **25** was fully consumed in the initial reactions, isolation of the product in sufficient purity for NMR identification was problematic, and mass spectrometry of the crude product was also inconclusive. No product formation was observed on treatment of the crude reaction mixtures with glycine *t*-butyl ester as above. Similar problems with such benzyl halides have been previously reported.[193]

We also attempted the conversion of alcohol **25** to the corresponding benzyl triflate, by treatment with triflic anhydride and pyridine, and to the benzyl tosylate, by treatment with tosyl chloride and pyridine, but these reactions were also unsuccessful as the products again proved unstable and impossible to isolate.



Scheme 35: formation of 2-bromo-2-(2,4-dimethoxyphenyl) (*S*-paramethoxybenzyl) ethanethiol and benzyl sulfonates and attempted displacements.

The low stability of these benzyl halides and sulfonates is likely to be due to favourable elimation of the halide or sulfonate group, due to stabilisation of the resulting

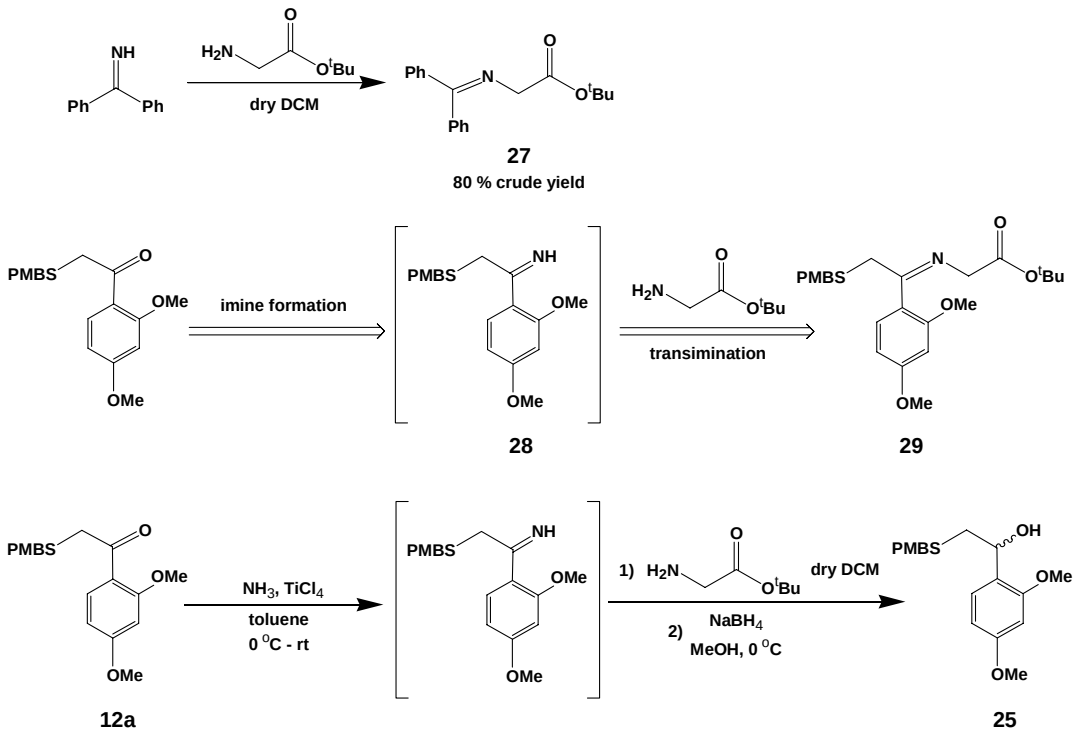
carbocation by the phenyl ring methoxy substituents, or possibly by intramolecular nucleophilic attack of the sulfur atom to form a three-membered ring sulfonium cation and subsequent elimination of the PMB group to form the 2,4-dimethoxyphenyl thiirane.

A further possible route to the desired auxiliary-amino acids was also investigated. Benzophenone Schiff base derivatives of amino acids, such as **27**, can be formed via transimination by simply adding benzophenone imine to the appropriate amino acid derivative (see **Scheme 36**).[194] We envisaged that conversion of ketone **12a** to the corresponding imine **28**, followed by addition of glycine *t*-butyl ester, might therefore result in the formation of imine intermediate **29**, which could then be reduced to the desired product.

As a trial of this approach, benzophenone glycine *t*-butyl ester **27** was formed as shown in a crude yield of 80 %. Although TLC analysis indicated complete reaction and **27** was detected by mass spectrometry of the crude product following work up, only a small amount (2 %) of the product was isolated by column chromatography. The major product isolated from the column was in fact benzophenone, arising from decomposition of the imine product during chromatography.

Despite this problem, we envisaged that this approach could still be successful if chromatographic purification was avoided until the final step. Ketone **12a** was thus treated with ammonia in the presence of titanium tetrachloride. TLC and mass spectrometry analysis of the crude product of this reaction following work up were

inconclusive. Upon addition of glycine *t*-butyl ester to a solution of the crude product however, evolution of gas was observed, consistent with imine formation in the first step. TLC analysis of the reaction was again inconclusive, but the transiminated product **29** was detected by mass spectrometry of the crude product following work up. Treatment of the crude product with sodium borohydride however, afforded only alcohol **25**. This suggests that, although formation of imine intermediate **29** appeared to occur to some extent, it had rapidly broken down to regenerate ketone **12a**, despite the avoidance of chromatography.

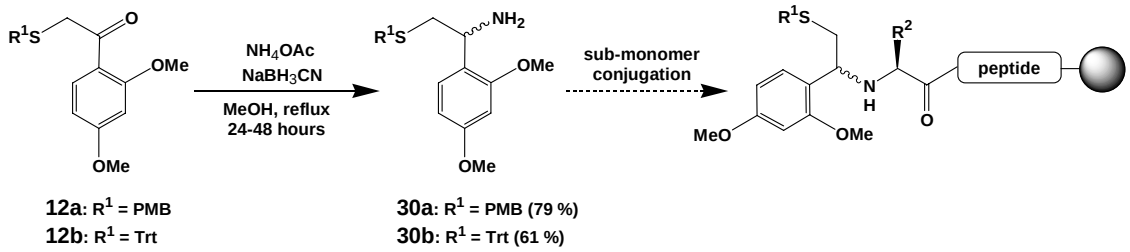


Scheme 36: transimination strategy.

At this stage we decided to revisit our original reductive amination approach, as both the intermediate benzyl halides for the displacement approach and the benzylimine for

the transimination approach appeared too unstable to be practically useful. Successful conditions for this transformation were eventually developed, based upon a method used for similar substrates by Williams and co-workers.[195]

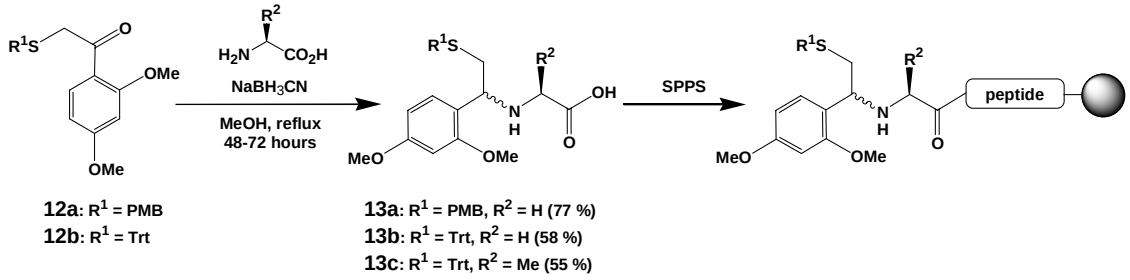
Treatment of ketones **12a** and **12b** with ammonium acetate and sodium cyanoborohydride in refluxing methanol for 24 – 48 hours afforded the *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) amines **30a** and **30b**, which can be used for the formation of auxiliary-peptides via sub-monomer conjugation (see **Scheme 37**).



Scheme 37: formation of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) amines **30a** and **30b**.

PMB-protected benzylamine **30a** could be easily isolated by acidification of the reaction mixture and washing with Et_2O to remove unreacted **12a**. Adjustment of the pH of the aqueous fraction to pH ≈ 10 with powdered potassium hydroxide, followed by extraction with DCM, afforded **30a** in good yield and purity. This acid-base extraction procedure was not possible for **30b** due to the acid-lability of the trityl protecting group, and therefore it was necessary to purify this product by column chromatography. The difficult isolation of the product from a close-running impurity contributed to the lower yield of **30b**.

These encouraging results prompted us to attempt formation of our preferred target auxiliary-amino acid cassettes under the same conditions. Gratifyingly, unprotected glycine and alanine were sufficiently soluble under the reaction conditions for successful reductive amination to occur in 77 – 55 % yield to afford auxiliary amino acid cassettes **13a** – **13c**, ready for incorporation into a peptide via standard SPPS (see **Scheme 38**).



12a: R^1 = PMB
12b: R^1 = Trt

13a: R^1 = PMB, R^2 = H (77 %)
13b: R^1 = Trt, R^2 = H (58 %)
13c: R^1 = Trt, R^2 = Me (55 %)

Scheme 38: synthesis of auxiliary-amino acid cassettes **13a** – **13c**.

Slightly longer reaction times were required than for amination with ammonium acetate. The addition of a small amount of DCM to the reaction mixture was found to aid formation of trityl-protected cassettes **13b** and **13c** by increasing the solubility of the starting ketone **12b**, although a lower yield for these products relative to the PMB species was observed as seen for amine **30b**. The use of unprotected amino acids enabled formation of cassettes **13a** – **13c** in a single step, without the further deprotection step required for the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliaries. The high polarity of **13a** – **13c** did cause some difficulties during chromatographic purification however, which led to a slight decrease in yield in each case. Auxiliary-alanine cassette **13c** was isolated as a 3:2 mixture of diastereomers.

Based upon the discovery that Fmoc protection was not necessary for the amino group of the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary-alanine cassette **11c**, we reasoned that Fmoc protection would also not be required for glycine cassettes **13a** and **13b**, where the amino group is similarly adjacent to a tertiary carbon centre, and alanine cassette **13c**, where the amine group lies between two tertiary carbons.

PMB and trityl protected auxiliary-amino acid cassettes **13a** – **13c** ready for use in SPPS were thus synthesised in only two steps in good overall yields (53 % - 75 %) utilising direct reductive amination of initially unreactive ketones **12a** and **12b** with the unprotected amino acid. This synthesis again represents a considerable improvement over the original route, being shorter and higher yielding and compatible with sensitive peptide modifications due to the use of the PMB and trityl protecting groups. Auxiliary benzylamines **30a** and **30b** for use in sub-monomer strategies could also be formed in good yield by this route.

A similar strategy for the assembly of the *N*α-(1-(4-methoxyphenyl)-2-mercaptoethyl) auxiliary **5** via reductive amination of the corresponding ketone with an amino acid to form an auxiliary-amino acid cassette has since been used by Botti and co-workers.[133] This synthesis utilised the *para*-methylbenzyl protecting group however, and involved a less straightforward two-step amination reduction procedure. Use of amino acid ethyl esters for the reductive amination also necessitated a further deprotection step. This group also noted the difficulty of Fmoc protection of the auxiliary-amino acid cassette and the possibility of successful coupling of the unprotected cassette, although careful

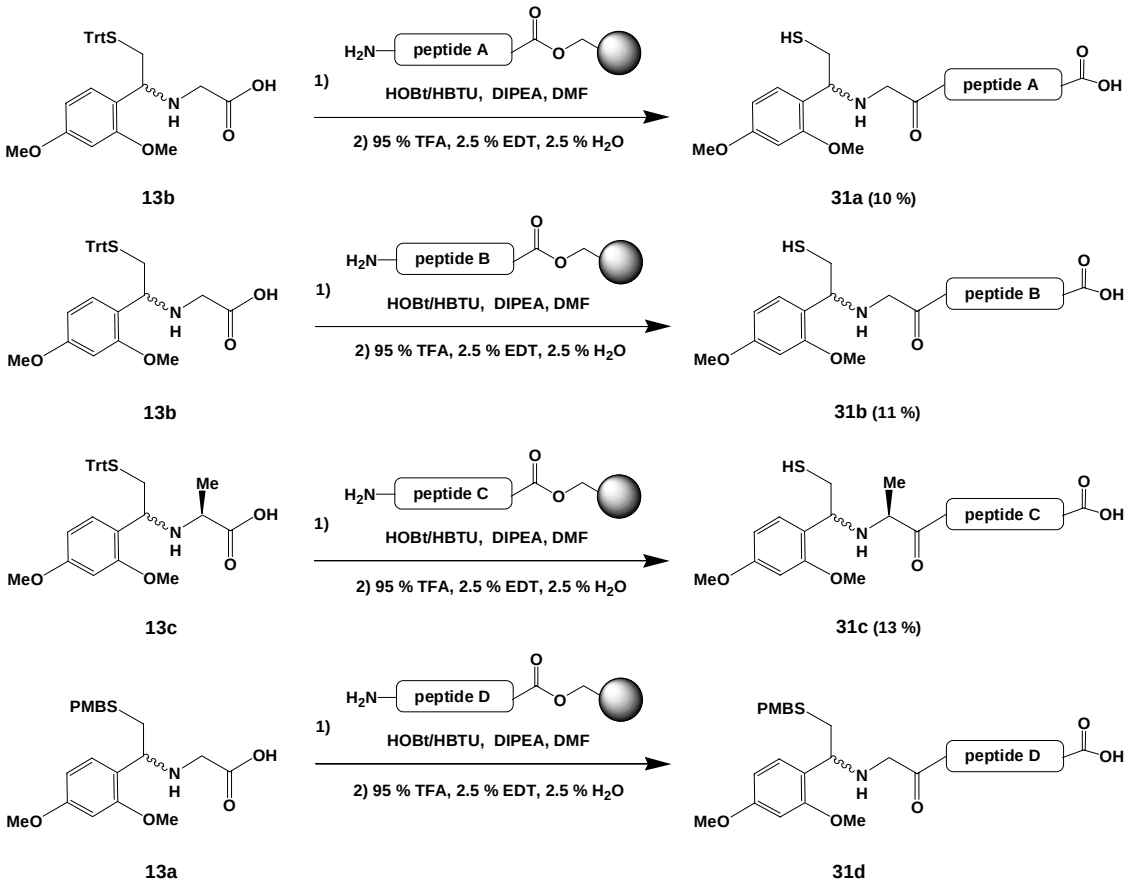
control of the reaction appeared to be necessary to avoid side reactions under the coupling conditions used.

1.2.1 Synthesis of auxiliary-peptides

Peptides **31a** – **31c** and glycopeptide **31d** bearing the *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary were synthesised on NovaSyn®TGT resin preloaded with serine as described for auxiliary **4**. Synthesis was conducted on a 0.05 mmol scale by manual Fmoc SPPS, using 10 equivalents of each Fmoc amino acid per coupling and HBTU/HOBt coupling reagents in the presence of Hünigs base. Fmoc removal was carried out with 20 % v/v piperidine in DMF. Glycopeptide **31d** was prepared using the Fmoc-Ser[$(OAc)_3$GalNAc]-OH cassette, prepared and incorporated as described in reference 196. Synthesis was monitored by the Kaiser ninhydrin test[190] and LC-MS analysis of samples of the peptide.

Auxiliary-amino acid cassettes **13a** – **13c** were coupled to peptides **A** - **C** and glycopeptide **D**, corresponding to fragments of GlyCAM-1, using five equivalents of amino acid and coupling reagents (see **Scheme 39**). The couplings of both auxiliary-glycine cassettes **13a** and **13b** and auxiliary-alanine cassette **13c** went to completion as monitored by LC-MS, with no evidence of coupling of auxiliary dipeptides or other byproducts resulting from acylation of the secondary amine of **13a** – **13c**. This result is in contrast with the suggestion by Botti and co-workers that this coupling step must be

carefully controlled,[133] and confirms that Fmoc protection is not required for these species.



Scheme 39: synthesis of auxiliary peptides **31a - 31c** and PMB-protected auxiliary glycopeptide **31d**.

Following coupling of cassettes **13a - 13c**, the resulting peptides were cleaved from the resin with concomitant side chain deprotection by treatment with 95 % v/v TFA, with 2.5 % v/v ethanedithiol (EDT) as a scavenger reagent. The trityl thiol protecting group

was also cleaved in this step (see **Scheme 39**). The crude peptides were then purified by semi-preparative reverse-phase HPLC to afford auxiliary-peptides **31a** – **31c** in 10 % - 13 % overall yield as shown and PMB-protected auxiliary-glycopeptide **31d**. The ESI mass spectrum for purified auxiliary-peptide **31b** is shown in **Figure 30**.

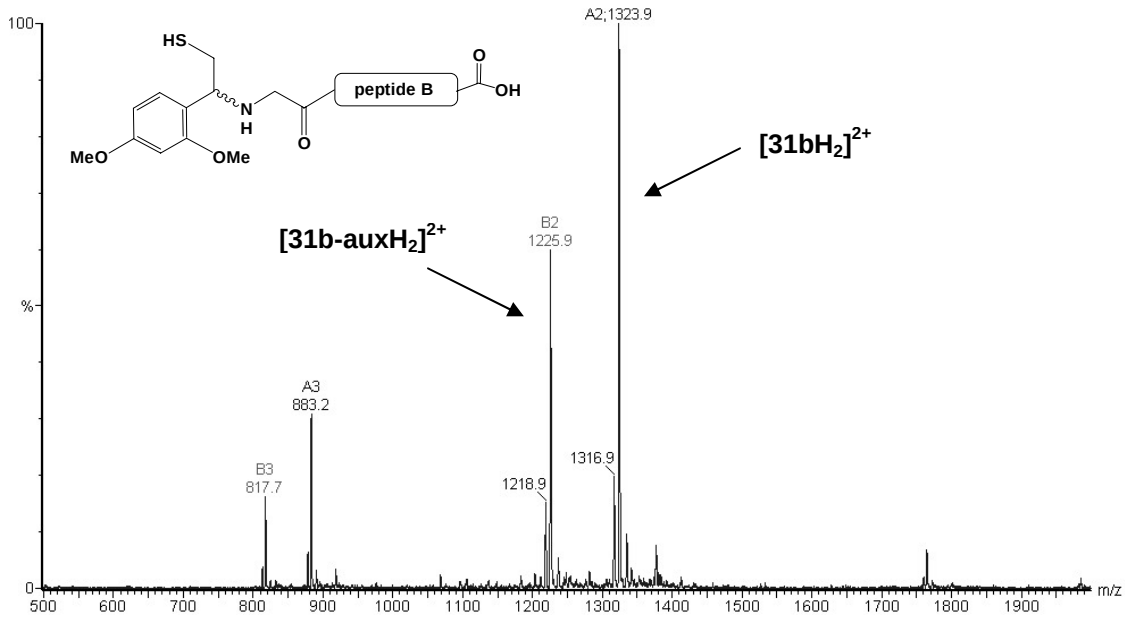
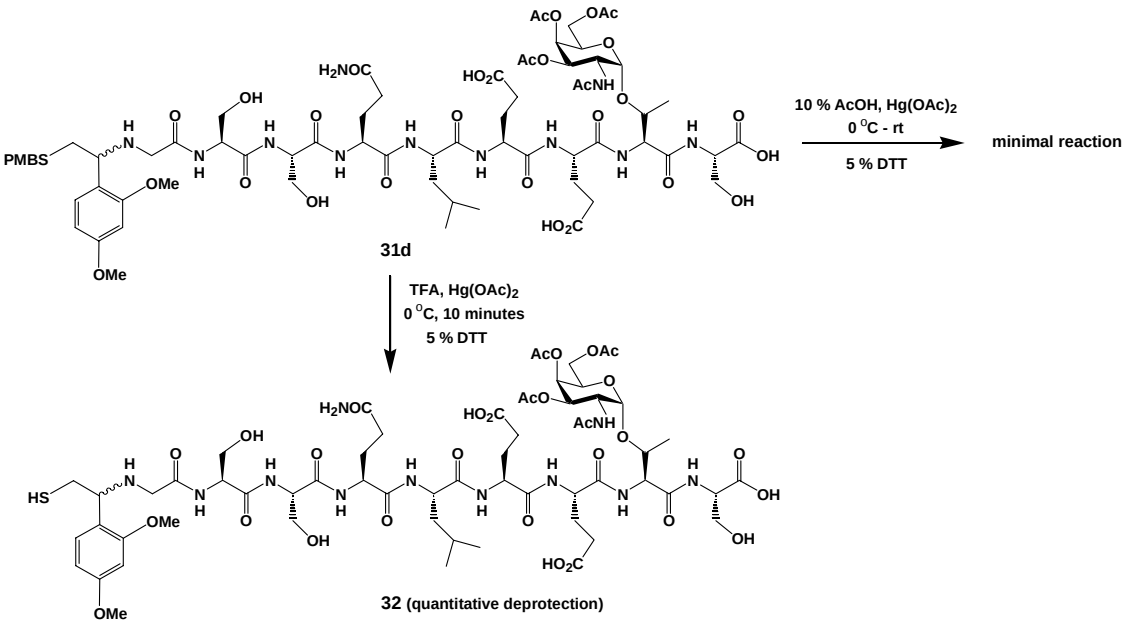


Figure 30: ESI positive ion mass spectrum of purified **31b**, showing characteristic fragmentation of auxiliary at the benzylic position (calculated mass = 2646.0 Da, observed: 1323.9 Da $[MH_2]^{2+}$, 1225.9 Da $[M\text{-}AuxH_2]^{2+}$, 883.2 Da $[MH_3]^{3+}$, 817.7 Da $[M\text{-}AuxH_3]^{3+}$). peptide B = KIIEGFVTGAEDIISGASRITKS.

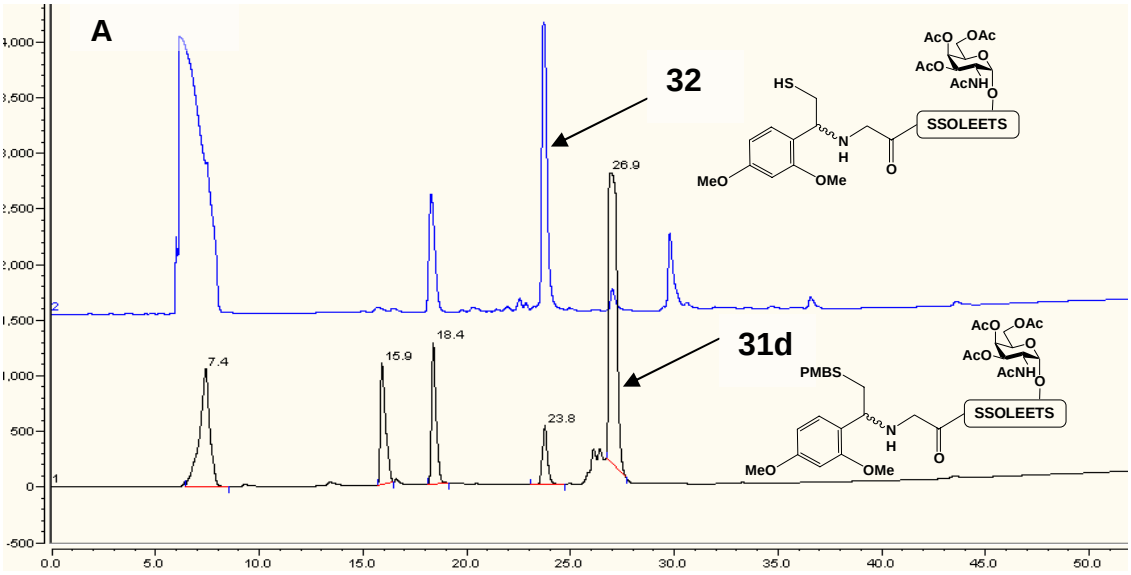
Removal of the PMB protecting group from glycopeptide **31d** was initially attempted under the conditions which had been successful for deprotection of 4,5,6-trimethoxy-2-mercaptobenzyl-peptides **23a** – **23c** (see **Scheme 40**). Treatment of **31d** with mercury(II)acetate in aqueous acetic acid, followed by the addition of DTT, had little

effect in this case however and only a minimal amount of the deprotected product was formed (black trace, **Figure 31**). The use of neat TFA at 0 ^{o}C instead of aqueous acetic acid however, led to almost complete deprotection within 10 minutes as shown by the blue trace in **Figure 31**.



Scheme 40: removal of PMB protection from auxiliary-glycine glycopeptide **31d**.

In summary, *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)-peptides could be efficiently assembled using either trityl- or PMB-protected building blocks, which could be synthesised in good yield in only two steps. Fmoc protection was not required for alanine or glycine building blocks. The trityl group was cleaved concomitantly with release from the solid phase to afford auxiliary-peptides **31a** – **31c** ready for use in cysteine-free peptide ligation. Adapted deprotection conditions for the removal of the PMB protecting group were applied to glycopeptide **31d** to afford auxiliary-glycopeptide **32**, similarly ready for use in ligation.



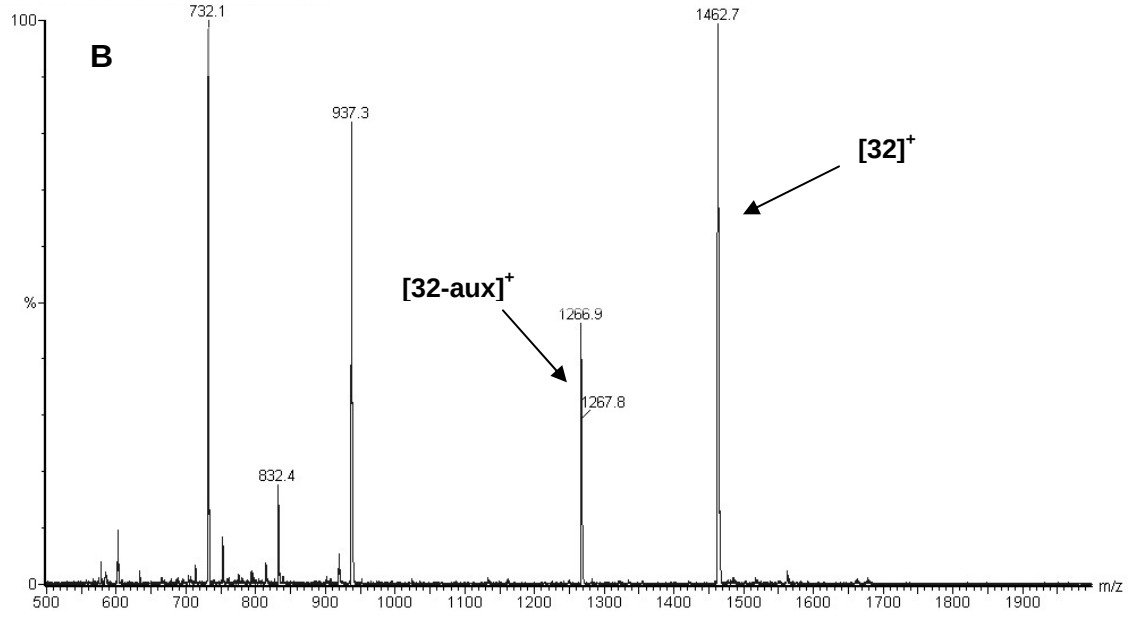


Figure 31: deprotection of auxiliary glycopeptide **31d**. **A**: HPLC purification UV traces for PMB deprotection of **31d**. Black trace: 10 % v/v AcOH, $Hg(OAc)_2$, 0 °C – RT, 1 hour, then 5 % w/v DTT, 1 hour; blue trace: neat TFA, $Hg(OAc)_2$, 0 °C, 10 minutes, then 10 % v/v TFA, 5 % w/v DTT, 1 hour. **B**: ESI positive ion mass spectrum of purified **32** (calculated mass = 1462.5, observed: 1462.7 $[M]^+$, 1266.9 $[M\text{-}Aux]^+$). A fragment peak (937.3) due to loss of auxiliary and peracetylated GalNAc is also seen.

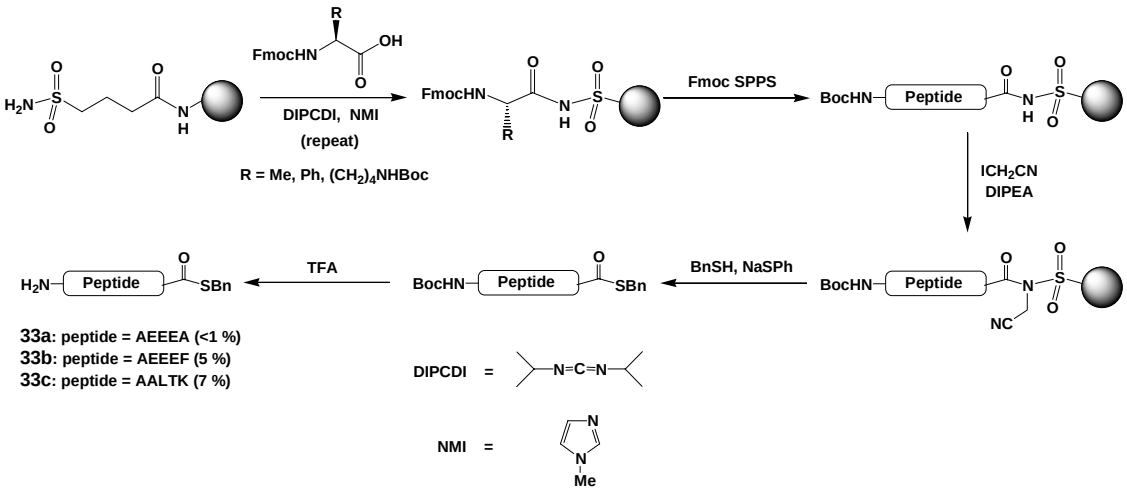
2. Synthesis of thioester peptides

Following the successful synthesis of the auxiliaries and their efficient incorporation into peptides, short thioester peptides were required to test the auxiliary-mediated ligation reaction for a range of ligation junctions. The sulfonamide "safety-catch" method was chosen for the synthesis of the thioester peptides due to previous experience of this approach in our group.[85,118,196] The adoption of this approach also allowed us to investigate some of the recent innovations in this area, specifically the use of the "double-linker" methodology[62] and the formation of thioesters via intramolecular *N* to *S* acyl shift[173] (see introduction, **section 2.3.10**).

2.1 Sulfonamide safety catch peptides: initial coupling with DIPCDI

The target thioester peptides had the sequences GLRG, AEEEA, AEEEF and AALTK. These were based on fragments of GlyCAM-1 and were designed to allow a range of ligation junctions to be examined. The initial synthesis of the *C*-terminal alanine, phenylalanine and lysine peptides was conducted on a 0.05 mmol scale on 4-sulfamylbutyryl safety catch resin with a loading of 1.1 mmol/g. Coupling of the *C*-terminal amino acid was carried out in DCM/DMF (4:1 v/v) with diisopropylcarbodiimide (DIPCDI) and 1-methyl imidazole (NMI), using four equivalents of each reagent and the appropriate Fmoc-protected amino acid. (see **Scheme 41**) This coupling was repeated as the initial coupling can be inefficient for safety catch resins.[170,171]

Elaboration of the peptide was then conducted by standard SPPS using HOBt/HBTU coupling with five equivalents of each Fmoc-amino acid exactly as described for the synthesis of the auxiliary-peptides. In each case coupling of the *N*-terminal residue was carried out with the Boc-protected amino acid in order to avoid the requirement for an additional Fmoc deprotection step.



Scheme 41: synthesis of peptide thioesters **33a** – **33c** with initial coupling with DIPCDI.

Following completion of the solid phase peptides, activation of the sulfonamide group was carried out with excess iodoacetonitrile in the presence of Hünigs base for 24 hours with the exclusion of light. Thioester formation with concomitant release from the resin was then carried out by treatment with excess benzyl mercaptan and catalytic sodium thiophenoxide for a further 18 – 24 hours.[108] The resin was then removed by filtration and the filtrate was concentrated and treated with 95 % v/v TFA, 2.5 % v/v EDT to

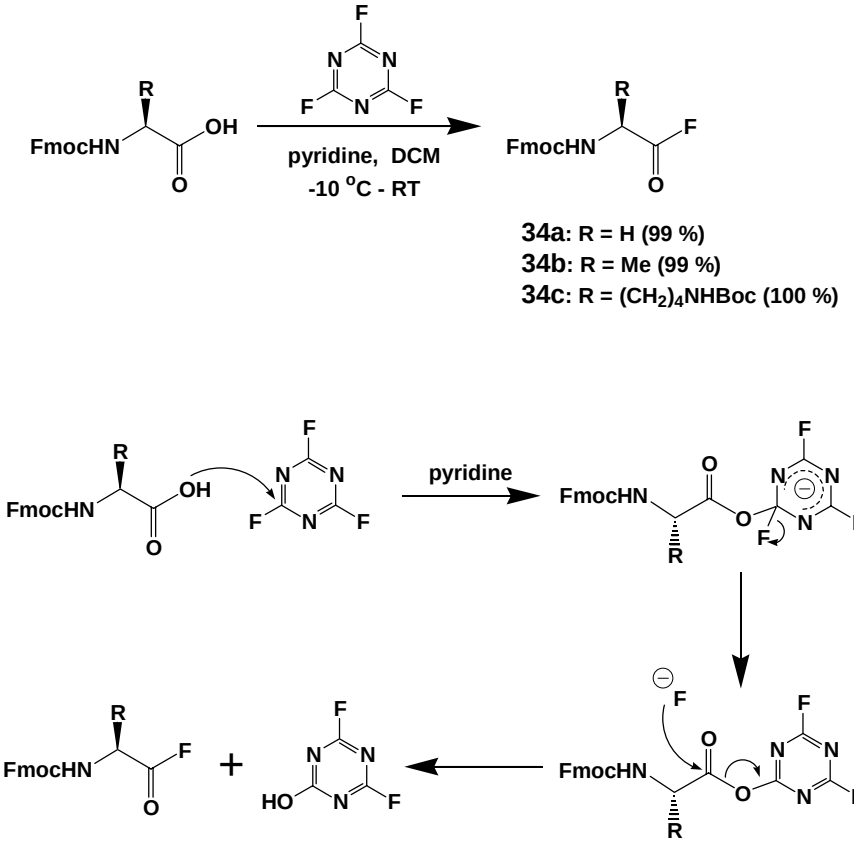
remove the side chain and *N*-terminal protecting groups. The crude products were then purified by semi-preparative HPLC.

Unfortunately HPLC purification afforded the target thioesters **33a** – **33c** in extremely low yields (< 1 % for alanine thioester **33a**, see **Scheme 41**). The crude products appeared to be relatively pure with no significant side products present, but product recovery was very low.

Due to the previously discussed problems of monitoring synthesis on sulfonamide safety catch resins (see introduction, section **2.3.10**) it was difficult to determine at which stage the synthesis had failed. The two most likely problem steps were the initial loading of the *C*-terminal amino acid and the final activation and cleavage step, as the remaining SPPS couplings and deprotections had gone to completion as monitored by the Kaiser ninhydrin test.

2.2 Sulfonamide safety catch peptides: initial coupling with acid fluoride

The very poor yields obtained for the peptide thioesters by the above approach led us to investigate the acid fluoride method for the coupling of the *C*-terminal amino acid.[171] Protected glycine, alanine and lysine were therefore treated with cyanuric fluoride in the presence of pyridine as shown in **Scheme 42** to afford the acid fluoride products **34a** – **34c** in quantitative or near-quantitative yield after work up, in sufficient purity to be used directly in SPPS.

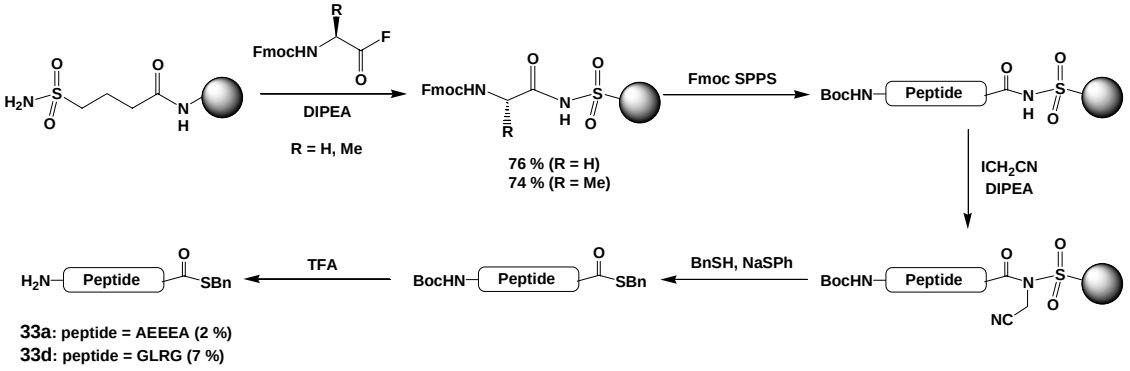


Scheme 42: synthesis of acid fluorides **34a** – **34c**.

Fmoc-protected glycine and alanine acid fluorides **34a** and **34b** were coupled to 4-sulfamylbutyryl safety catch resin in the presence of Hünigs base for two hours (see **Scheme 43**). UV Fmoc analysis of the resulting peptidyl resins showed a loading of 76 % for glycine and 74 % for alanine with respect to the given value for the resin loading. SPPS and the activation and cleavage and deprotection steps were then carried out as before to afford *C*-terminal alanine thioester **33a** and *C*-terminal glycine thioester **33d**.

Despite the efficient initial couplings however, *C*-terminal alanine thioester **33a** was again obtained in a low yield of 2 %. Although this represents an improvement over the previous yield obtained for **33a** via the DIPCDI coupling method, and sufficient

material for further studies was obtained, the low yield still represents a disappointing result. Glycine thioester **33d** was also obtained in a low yield of 7 %. These results suggested that the activation and cleavage reactions were the problem steps. No significant side products were seen during HPLC purification as before, suggesting that the activation and cleavage reactions were clean but inefficient, and that the majority of the peptide had remained on the resin.

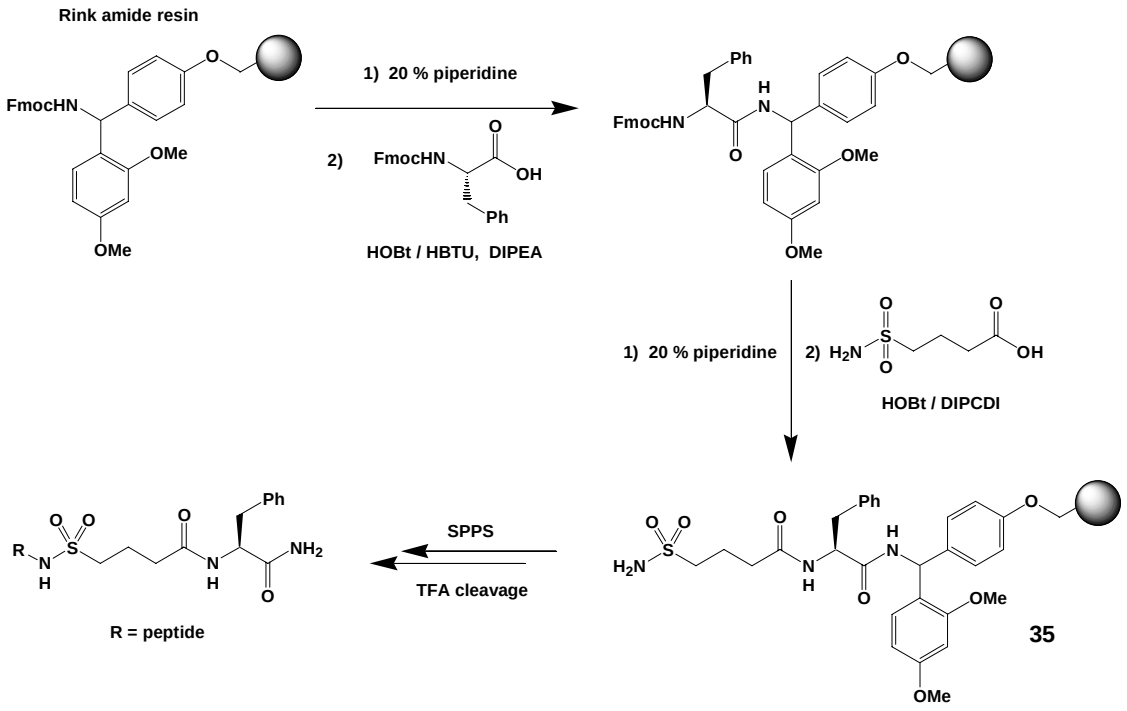


Scheme 43: synthesis of thioester peptides **33a** and **33d** via initial coupling with acid fluoride.

2.3 "Double linker" sulfonamide safety catch peptides

Examination of the activation and cleavage reactions for sulfonamide safety catch peptides is difficult due to the fact that samples for analysis must be prepared by the same two step procedure. To solve this problem, we sought to apply the recently-developed "double-linker" strategy[62] (see introduction, section **2.3.10**) to allow easier analysis of these reactions via facile acidic cleavage of the peptide from the resin.

Alkane sulfonamide resin **35** was therefore prepared on Rink amide resin as shown in **Scheme 44**. Fmoc protected phenylalanine was coupled to Rink amide resin (loading 0.62 mmol/g) as a spacer, under standard HOBt/HBTU coupling conditions. The phenylalanine also functions as an aid to purification and characterisation by increasing the size, hydrophobicity and UV activity of the acid-cleaved peptide. Following Fmoc deprotection, 3-carboxypropanesulfonamide was added by HOBt/DIPCDI coupling, using three equivalents of each reagent, to form the double-linker sulfonamide resin **35**. SPPS and the activation and cleavage reactions can thus be easily monitored by acidic cleavage of the peptide at the phenylalanine residue as shown.



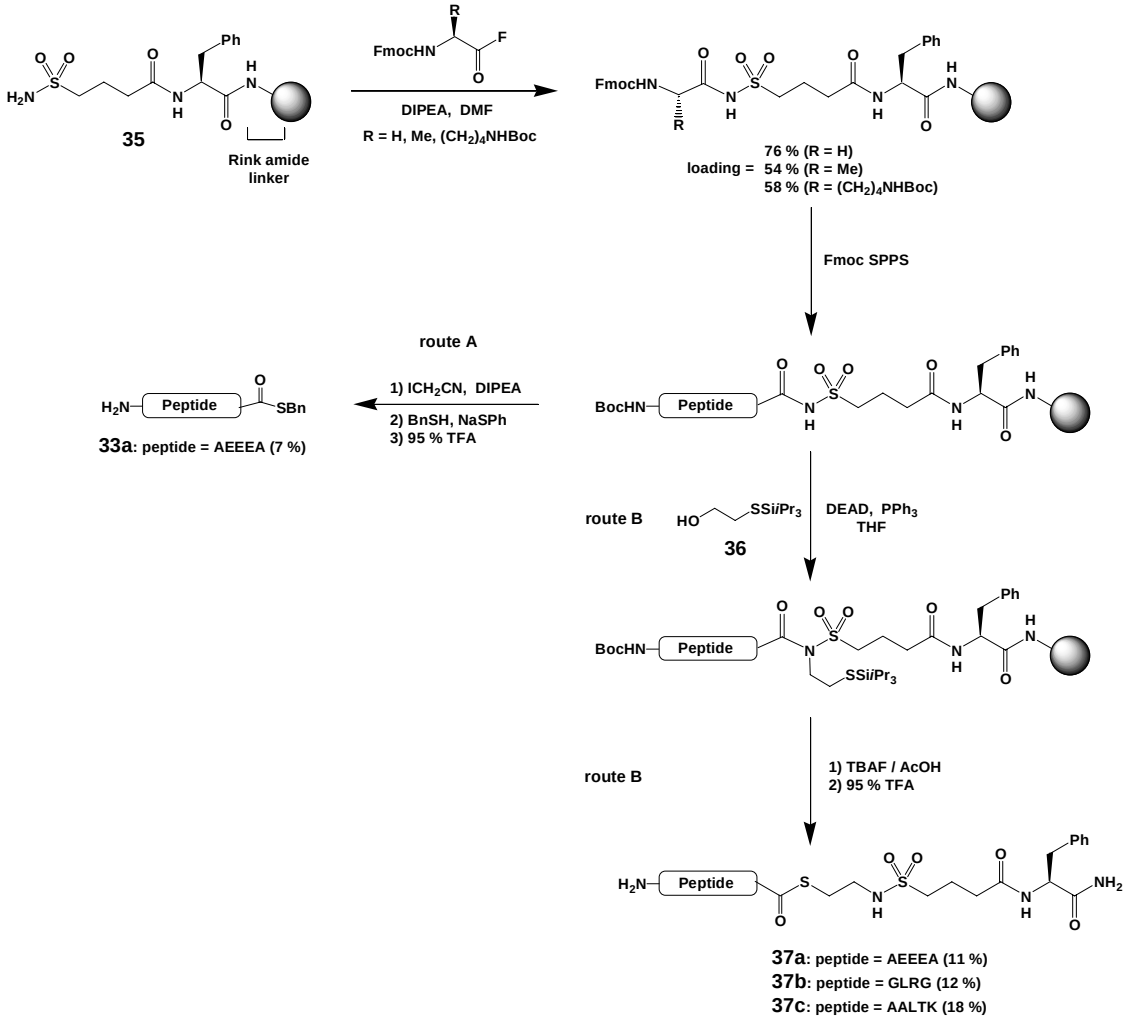
Scheme 44: synthesis of double-linker sulfonamide resin **35**.

The *C*-terminal amino acid residue of each of the thioester peptides AEEEA, AALTK and GLRG, was coupled to the sulfonamide resin **35** on a 0.1 mmol scale as the Fmoc-

protected acid fluoride in the presence of Hunigs base as before (see **Scheme 45**). UV Fmoc analysis showed a loading of 76 % for glycine and a loading of 55 % and 58 % for alanine and lysine respectively. Although these values are relatively low, they represent a loading sufficient to obtain the material required for ligation studies. These low loadings also indicate that acylation of the sulfonamide is difficult and hence side products arising from acylation of the remaining free sulfonamide by activated esters later in SPPS should not be formed to any significant degree.

Following completion of the peptides, thioester formation was carried out in one of two ways: by activation with iodoacetonitrile followed by cleavage with benzyl mercaptan to afford the benzyl thioester **33** (route A, **Scheme 45**), or activation by Mitsunobu alkylation with 2-hydroxy-(*S*-triisopropylsilyl)-ethanethiol **36**, followed by deprotection and spontaneous intramolecular $N - S$ acyl shift[173] (route B). In this second case the resin-bound alkyl thioester was subsequently released by TFA cleavage of the Rink amide linker to afford free thioester peptide **37**.

Following synthesis of the *C*-terminal alanine peptide (sequence AEEEA) on resin **35**, the peptide was cleaved from the resin as the benzyl thioester (**33a**) by activation with iodoacetonitrile followed by cleavage with benzyl mercaptan and subsequent TFA deprotection (route A, **Scheme 45**), as previously carried out for peptides synthesised on 4-sulfamylbutyryl safety catch resin.



Scheme 45: synthesis of thioester peptides on double-linker resin **35**, via standard activation/cleavage to form benzyl thioesters (route A) or intramolecular formation of alkyl thioesters (route B).

Thioester **33a** was the only product observed by LC-MS of the crude reaction mixture (see **Figure 32**), but a low yield of 4 % was again obtained for **33a** following HPLC purification. In this case however, use of the double-linker approach allowed the examination of the remaining resin-bound species by treatment of a sample of the resin with TFA (see **Figure 33**).

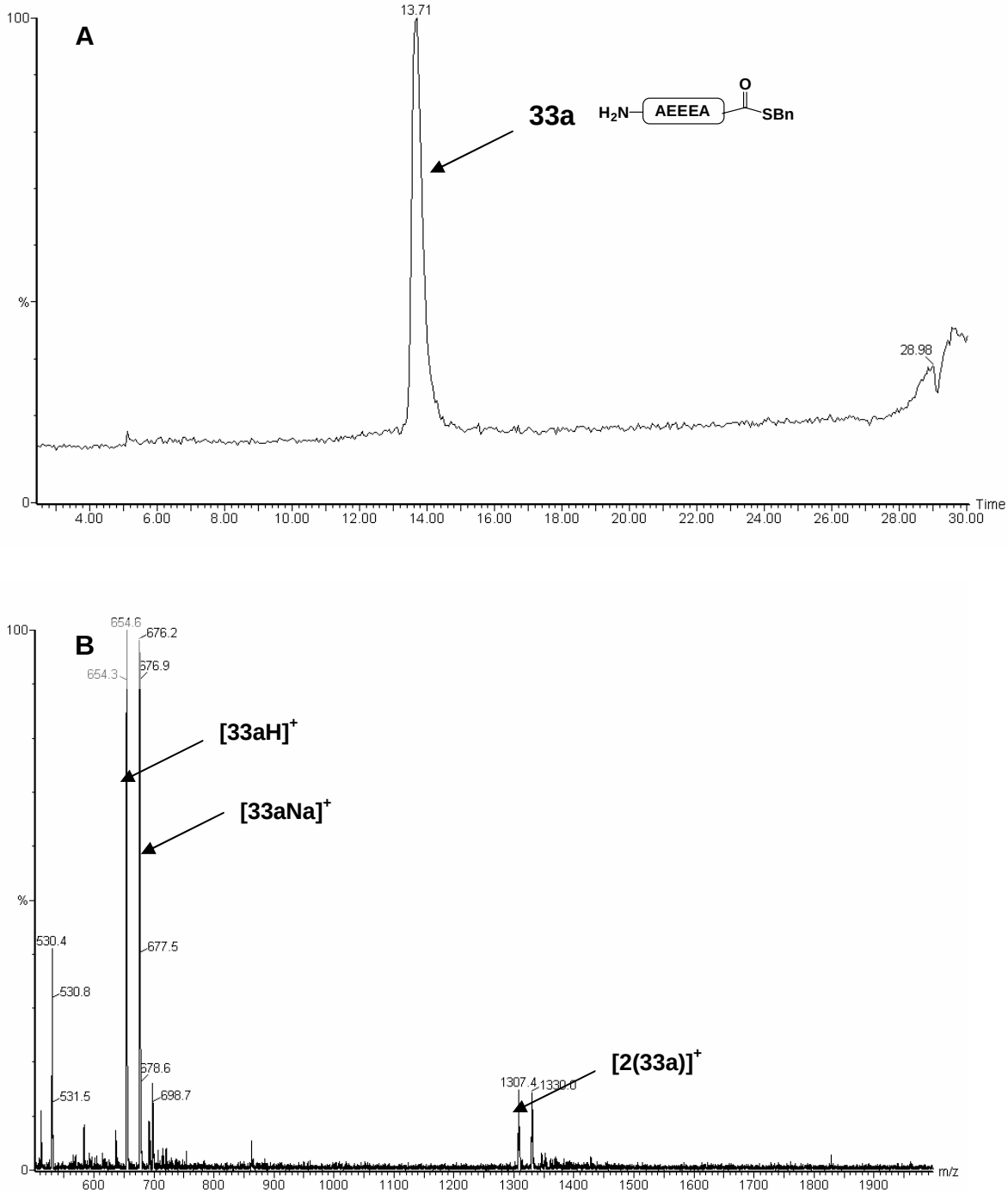


Figure 32: LC-MS analysis of crude reaction mixture for formation of thioester **33a**. **A**: total ion count (TIC) chromatogram; **B**: ESI spectrum of **33a** peak (calculated mass = 653.7 Da, observed: 654.6 Da $[MH]^+$, 676.2 Da $[MNa]^+$, 1307.4 Da $[2M]^+$).

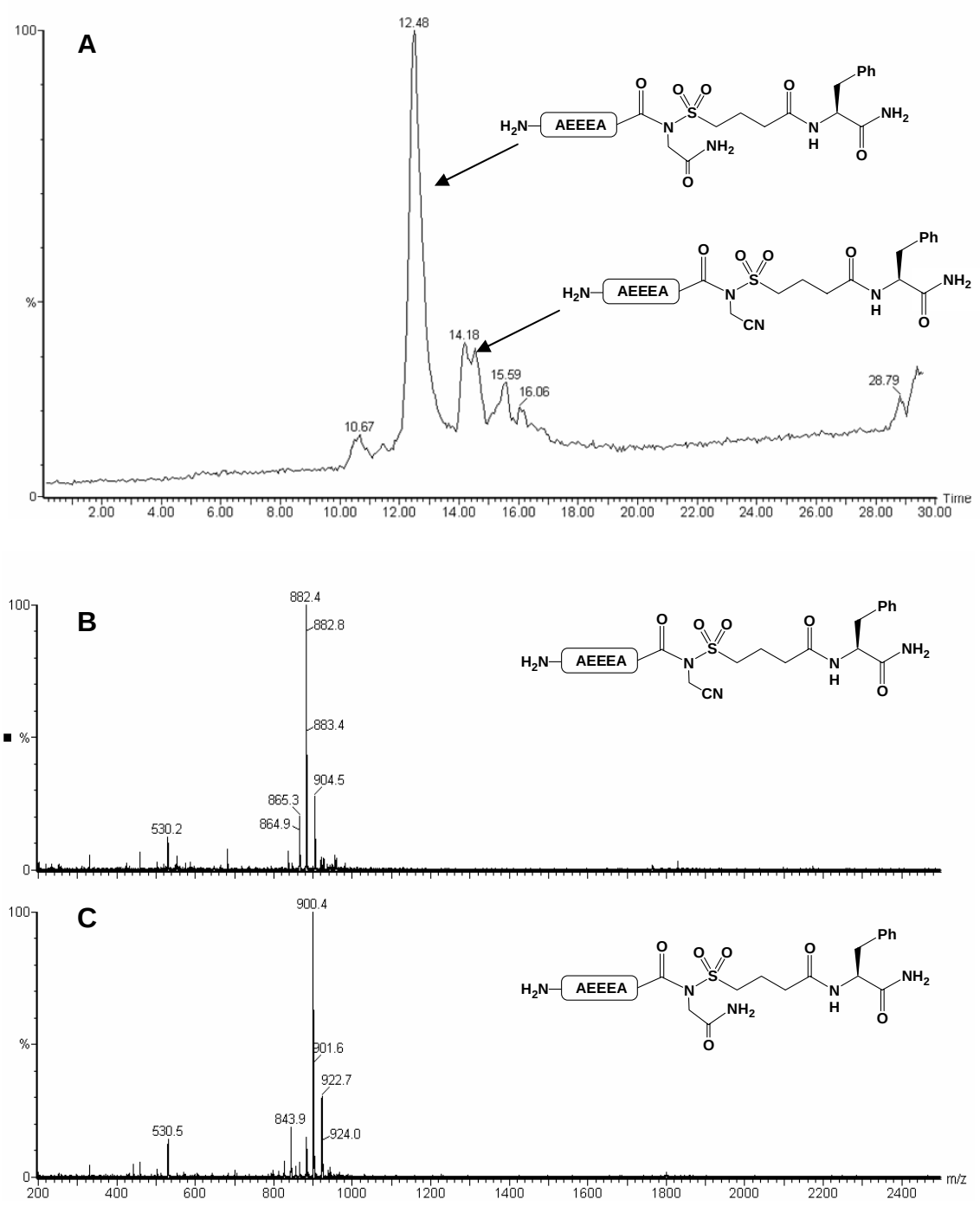


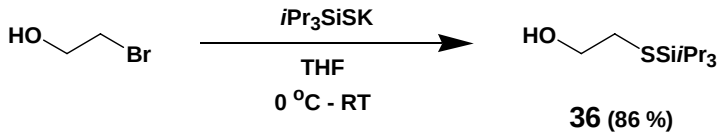
Figure 33: LC-MS analysis of resin-bound products. **A**: TIC trace showing alkylated sulfonamide peptides; **B**: ESI positive ion mass spectrum of minor peak (calculated mass = 881.9 Da, observed: 882.4 Da $[M]^+$); **C**: ESI positive ion mass spectrum of major peak (calculated mass = 899.9 Da, observed: 900.4 Da $[M]^+$).

LC-MS of the resulting crude product showed two major peaks due to products arising from incomplete reaction with benzyl mercaptan (see **Figure 33**). The smaller of these was the iodoacetonitrile-alkylated peptide. The larger peak corresponded to the peptide alkylated with acetamide (CH_2CONH_2) which is presumably formed by hydrolysis of the nitrile group under the aqueous resin cleavage conditions. Only a trace of non-alkylated peptide was present, and no shorter peptide side products arising from acylation of the free sulfonamide remaining from the initial acid fluoride coupling were observed. The alkylation step had therefore been efficient and the problem step in the thioester synthesis appeared to be the reaction of the activated peptide with benzyl mercaptan. The remaining resin was therefore resubjected to treatment with benzyl mercaptan and a further 3 % yield of thioester **33a** was isolated following HPLC purification.

Although the analysis of the activation/cleavage procedure allowed an improvement in the yield of thioester formation, resubjection of the release step was necessary and the amount of product recovered was still low. We therefore investigated the formation of thioesters via intramolecular $N - S$ acyl shift,[173] (route B, **Scheme 45**) as in this approach all peptide products are cleaved from the resin by the final treatment with TFA.

2-Hydroxy-(*S*-triisopropylsilyl)-ethanethiol **36** was prepared as shown in **Scheme 46**. The potassium salt of triisopropylsilyl (TIPS) thiolate was prepared by the addition of TIPS thiol to potassium hydride in pentane at 0 ^{o}C, and recrystallised from toluene to afford the salt as a white solid, which was used without further purification.

Displacement with bromoethanol by its addition to a suspension of the salt in THF afforded the product **36** in 86 % yield following purification by column chromatography.



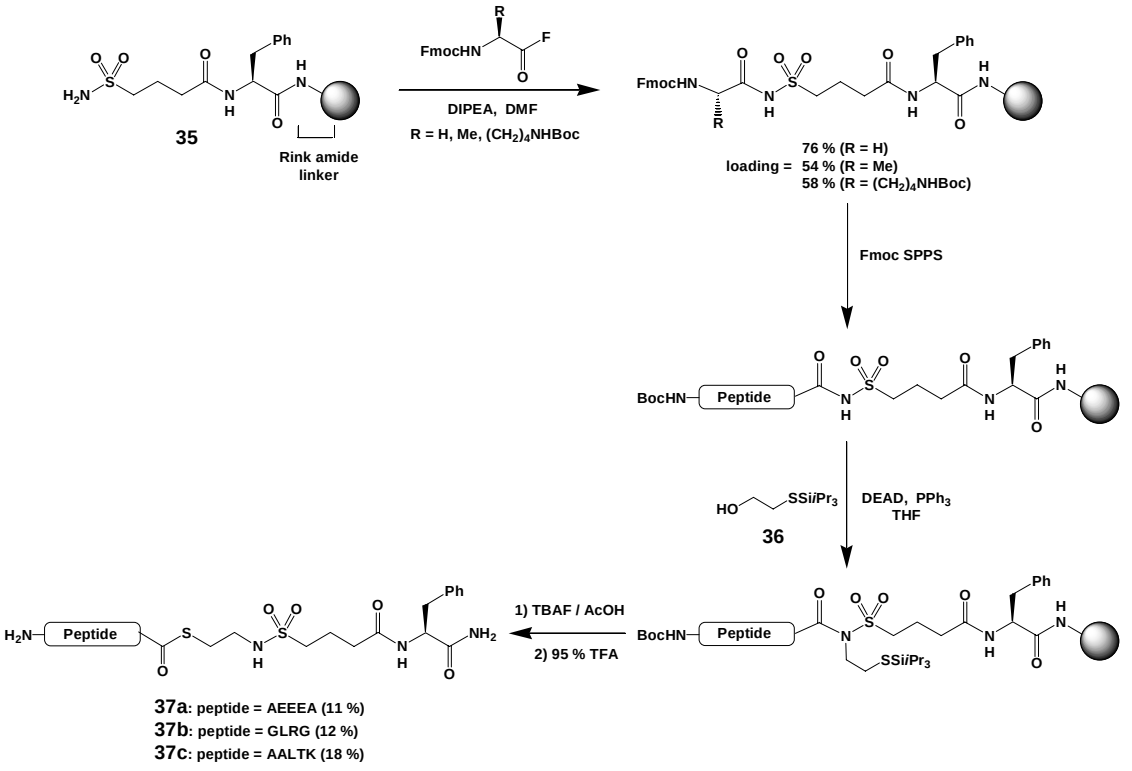
Scheme 46: preparation of **36**.

C-terminal alanine peptide AEEEA was again prepared on double-linker resin **35**, as in previous studies, and subjected to Mitsunobu alkylation with **36** with diethylazodicarboxylate (DEAD) and triphenylphosphine in THF (see **Scheme 47**).

Subsequent TFA treatment of a sample of the resin and MS analysis showed several peaks, one of which corresponded to the alkylated sulfonamide with fragmentation of the TIPS protecting group. The major peak corresponded to the non-alkylated starting material, however. Although the formation of complex mixtures of products arising from TFA cleavage of the alkylated peptide at this stage has been reported,[173] the presence of the starting peptide as the major peak suggested that alkylation was incomplete. The resin was thus resubjected to the Mitsunobu alkylation conditions and then taken on to the next step without further analysis.

The resulting peptide resin was treated with a solution of tetrabutylammonium fluoride (TBAF) and acetic acid in THF to remove the TIPS protecting group. The peptide was then cleaved from the resin with 95 % v/v TFA, 2.5 % v/v EDT and purified by HPLC.

The authors of the paper describing this approach state that intramolecular $N-S$ acyl shift to form the thioester occurs spontaneously upon removal of the TIPS group, but this rearrangement may actually occur during the TFA cleavage step. A species corresponding to the mass of the target thioester **37a** was isolated in 11 % yield, or 20 % with respect to the measured loading for the *C*-terminal alanine residue (see **Scheme 47**), along with a small amount of the non-alkylated peptide (not isolated). As the mass of the target thioester is identical to that of the deprotected but unrearranged acyl sulfonamide however, further tests were conducted to confirm the identity of the isolated product.



Scheme 47: synthesis of alkyl thioester peptides **37a - 37c** on double-linker resin **37**, via intramolecular thioester formation.

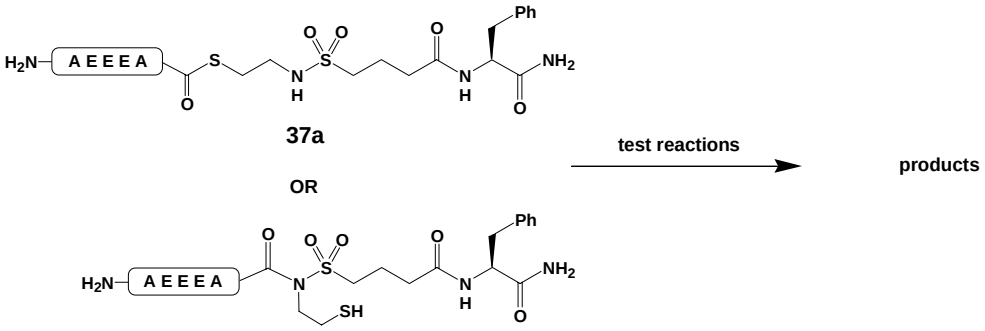
In previous studies, treatment of the products of NCL with nucleophiles [106,197] such as hydrazine[108,113] has been used to confirm that the rearrangement from the thioester to the amide has occurred, as the thioester reacts to form the hydrazide, whereas the rearranged amide product is unreactive. A sample of the isolated product was thus treated with 2 % v/v hydrazine hydrate in sodium phosphate buffer at pH = 8. LC-MS analysis of the reaction after three hours showed complete consumption of the starting material. Unfortunately the hydrazide product was not seen, presumably due to a very short retention time for this species. Instead a single major peak was observed corresponding to the mass of the thiol **38** released from the thioester product upon reaction (see **Figure 34**).

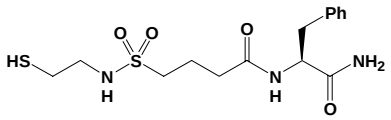
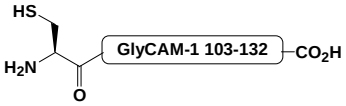
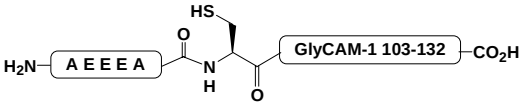
Treatment of a second sample of the isolated product with excess benzyl mercaptan led to formation of the benzyl thioester **33a** in under two hours. Thiol **38** was again observed, along with peaks corresponding to the oxidised disulfide. These results are consistent with the presence of thioester **37a**, but it is also possible that similar results might be observed from nucleophilic attack of benzyl mercaptan on the alkylated (and thus activated) but unrearranged acyl sulfonamide product.

Although such rapid reaction of the unrearranged acyl sulfonamide would be surprising given the slow nucleophilic cleavage observed for the activated sulfonamide resin, a further test was required to conclusively identify the isolated product as thioester **37a**.

As the unrearranged sulfonamide would not be reactive in cysteine ligation, a sample of the product was treated with an excess of an *N*-terminal cysteine peptide with the

sequence H-Cys-(GlyCAM-1 103-132)-OH as shown in **Figure 34**. After 24 hours, almost all of the starting material had been consumed, with formation of the expected ligation product. Thiol **38** was also observed. After 48 hours LC-MS indicated complete consumption of the starting material, which was thus identified as thioester **37a** containing no unrearranged acyl sulfonamide (see **Figures 35** and **36**).



conditions	products observed by LC-MS
2 % v/v hydrazine hydrate, 10 mM Na phosphate buffer pH = 8 3 hours	 38
benzyl mercaptan, 10 mM Na phosphate buffer pH = 8 2 hours	 33a + 38
 20 mM TCEP, 2 % MESNa 10 mM Na phosphate buffer pH = 8 48 hours	 + 38

GlyCAM-1 103-132 = TVEEELGKIIEGFVTGAEDIISGASRITKS

Figure 34: thioester test reactions.

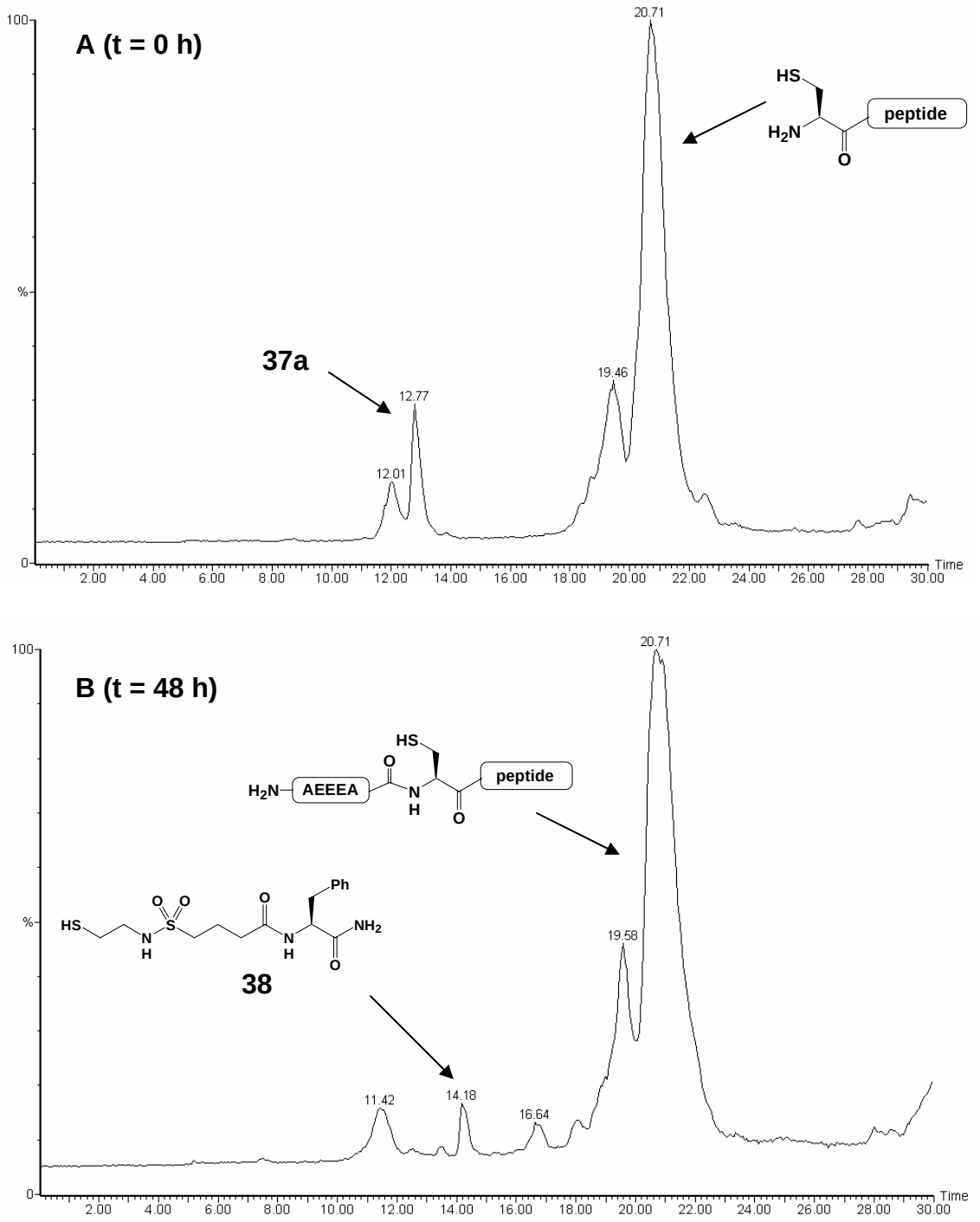


Figure 35: LC-MS analysis of thioester confirmation reaction with Cys-peptide. **A**: reaction at time = 0 hours, showing peaks for test product (assumed to be **37a**) and Cys-peptide; **B**: reaction at time = 48 hours, showing peaks for thiol **38**, ligation product and no remaining **37a**. peptide = TVEEELGKIIEGFVTGAEDIISGASRITKS.

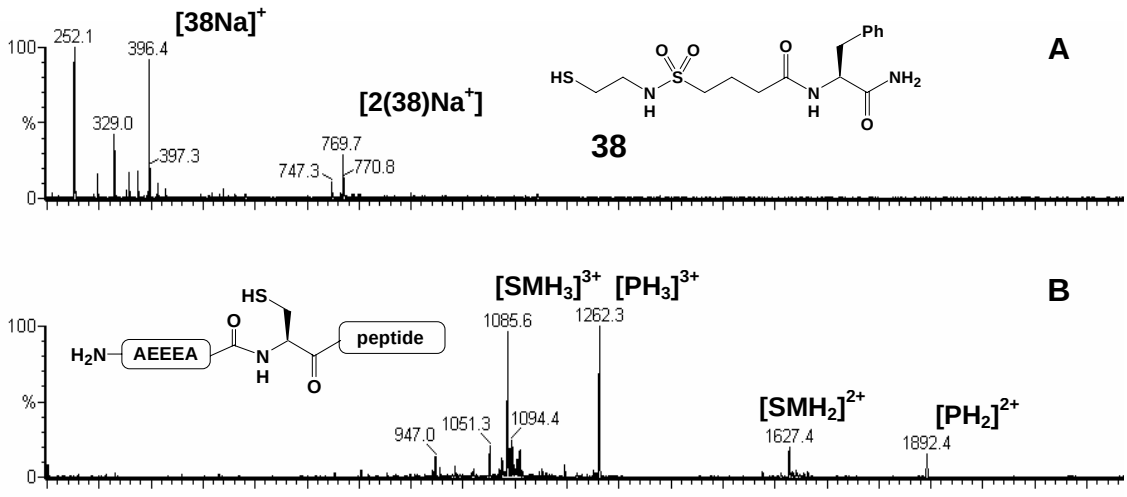


Figure 36: ESI positive ion mass spectra of ligation products from LC-MS **35B** above. **A**: mass spectrum of peak for thiol **38** (calculated mass = 373.5 Da, observed: 396.4 Da $[MNa]^+$, 769.7 Da $[2MNa]^+$); **B**: ESI positive ion mass spectrum of peak for ligation product (calculated mass = 3783.2 Da, observed: 1892.4 Da $[MH_2]^{2+}$, 1262.3 Da $[MH_3]^{3+}$), also showing signals for remaining Cys-peptide starting material (calculated mass = 3253.7 Da, observed: 1627.4 Da $[MH_2]^{2+}$, 1085.6 Da $[MH_3]^{3+}$). peptide = TVEEELGKIIEGFVTGAEDIISGASRITKS.

The intramolecular rearrangement approach was therefore found to be the most efficient method for the synthesis of the *C*-terminal alanine thioester, and *C*-terminal glycine thioester GLRG (**37b**) and *C*-terminal lysine thioester AALTK (**37c**) were subsequently synthesised by the same approach. In each case more than 50 % alkylation was achieved by subjection to the Mitsunobu conditions (see **Figure 37**). As this conversion provided sufficient material for ligation studies, the step was not repeated. The thiol deprotection and TFA cleavage steps were carried out exactly as before to afford thioesters **37b** and **37c** in 12 % and 18 % isolated yield respectively, after HPLC purification (see **Scheme 47**).

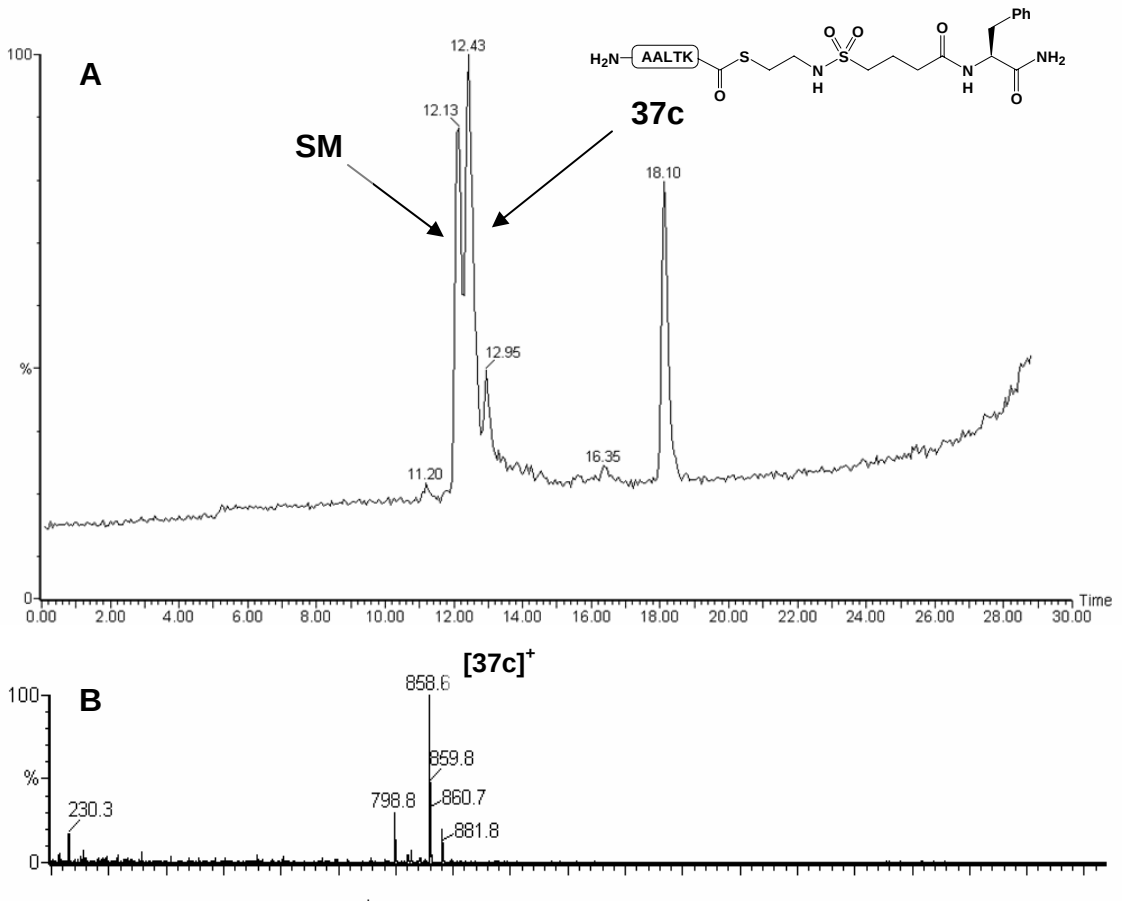


Figure 37: LC-MS analysis of crude reaction mixture for cleavage of *C*-terminal lysine thioester **37c** from the resin. **A**: TIC trace of crude reaction mixture, showing peaks for **37c** and for non-alkylated acyl sulfonamide (SM); **B**: ESI positive ion mass spectrum of **37c** peak (calculated mass = 858.1 Da, observed: 858.6 Da $[M]^+$).

3. Ligation studies

The auxiliary-peptides and short thioesters described in the previous sections were used to examine the scope and limitations of auxiliary-mediated ligations by conducting test ligations with auxiliaries **4** and **6** at a range of Xxx-Gly junctions and also at the Gly-Ala junction. Initial studies were conducted using the most straightforward junction, Gly-Gly. The compatibility of the ligation and auxiliary removal conditions with glycosylated peptides was also investigated at this stage. Studies of ligations at more sterically demanding junctions were then conducted.

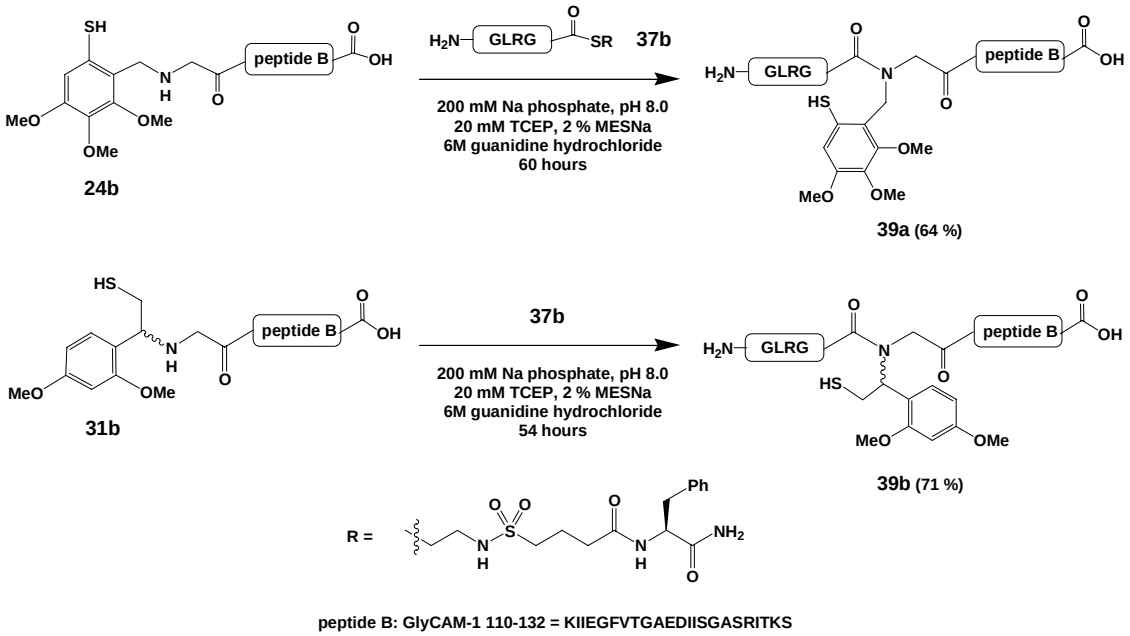
3.1 Ligation at Gly-Gly junctions

3.1.1 Ligation results

C-terminal glycine alkyl thioester **37b** was ligated with *N*-terminal auxiliary peptides **24b** and **31b** under typical denaturing NCL conditions, at pH = 8.0 with TCEP as the reducing agent and MESNa thiol additive as shown (**Scheme 48**). Both the thioester and auxiliary peptides were used in approximately 3 mM concentration. Reaction progress was monitored by LC-MS.

Ligation of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)-peptide **31b** with **37b** was virtually complete within 48 hours, and no further reaction was observed after 54 hours. Two different peaks corresponding to the mass of the product **39b** were observed by

LC-MS, corresponding to the two diastereomers of the product.[134,135] HPLC purification afforded the ligation product **39b** in 71 % yield (see **Figure 39**).



Scheme 48: Gly-Gly ligations with auxiliary-peptides **24b** and **31b**.

Ligation of 4,5,6-trimethoxy-2-mercaptobenzyl peptide **24b** proceeded slightly more slowly, and the reaction took 60 hours to reach completion. In this case unreacted **24b** eluted very close to the product **39a** from both LC-MS and HPLC columns, complicating analysis and purification somewhat.

Ligation product **39a** was isolated in 64 % yield by HPLC, containing traces of starting auxiliary-peptide **24b** (see **Figure 38**). In both ligations alkyl thioester **37b** was only observed at the early stages of each reaction, being rapidly converted to the MESNa thioester thereafter.

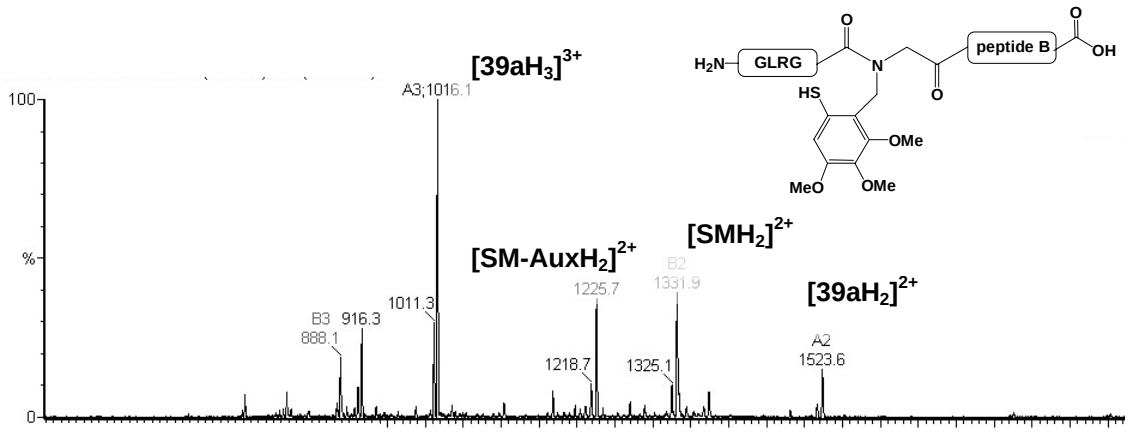


Figure 38: ESI positive ion mass spectrum of ligation product **39a** (calculated mass = 3045.5 Da, observed: 1523.6 Da $[MH_2]^{2+}$, 1016.1 Da $[MH_3]^{3+}$) containing traces of starting auxiliary-peptide **24b**. peptide B = KIIEGFVTGAEDIISGASRITKS.

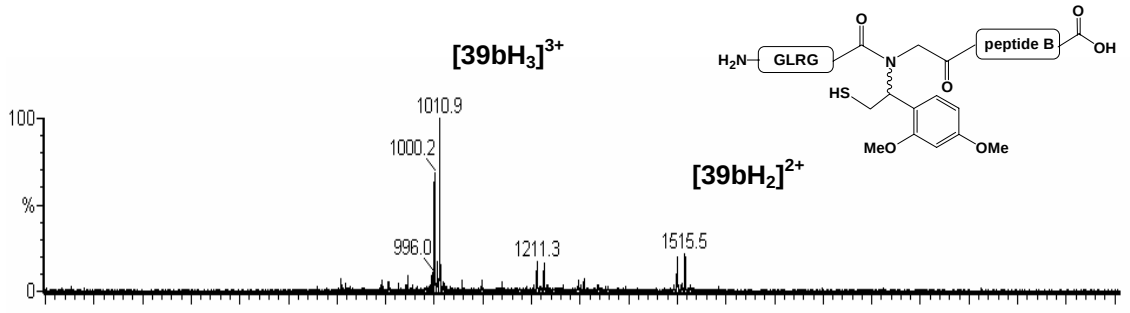
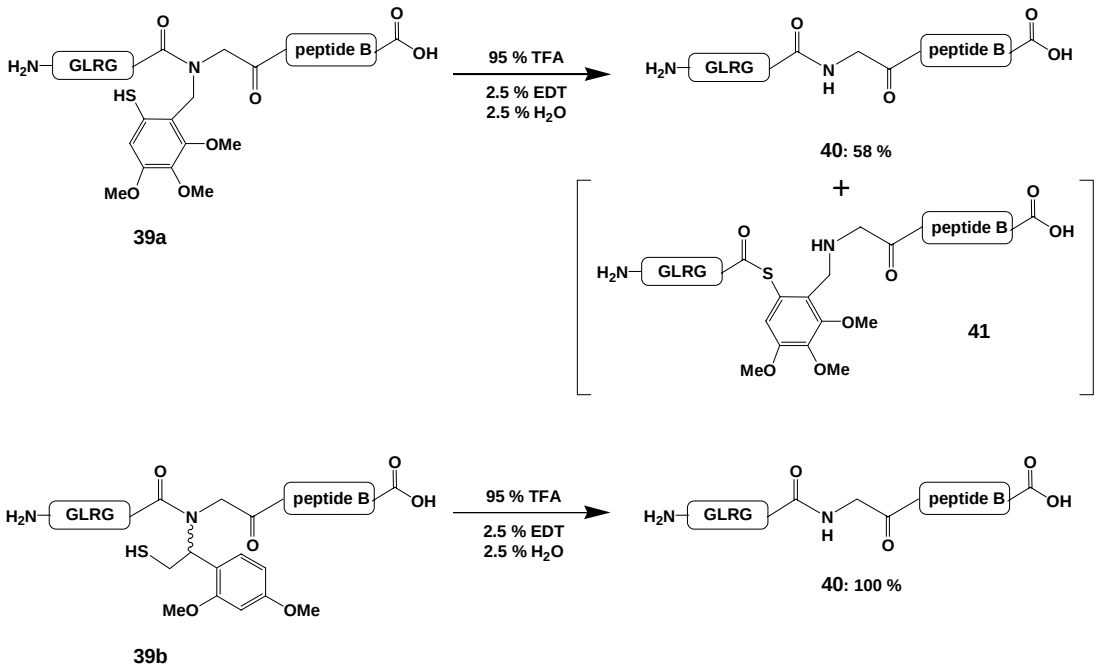


Figure 39: ESI positive ion mass spectrum of ligation product **39b** (calculated mass = 3029.5 Da, observed: 1515.5 Da $[MH_2]^{2+}$, 1010.9 Da $[MH_3]^{3+}$). peptide B = KIIEGFVTGAEDIISGASRITKS.

3.1.2 Auxiliary removal

Following successful formation and isolation of ligation products **39a** and **39b**, these peptides were treated with 95 % v/v TFA, 2.5 % v/v EDT for three hours to remove the auxiliary groups (see **Scheme 49**).



Scheme 49: auxiliary removal from ligation products **39a** and **39b**.

The *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** was efficiently removed from peptide **39b** under these conditions, to afford the desired native peptide **40** in quantitative yield following HPLC purification (see **Figure 40**). The successful auxiliary cleavage reaction confirmed that the key rearrangement of the thioester

intermediate to the amide-linked ligation product had taken place, as the unacylated auxiliary benzylamine is stable to TFA.

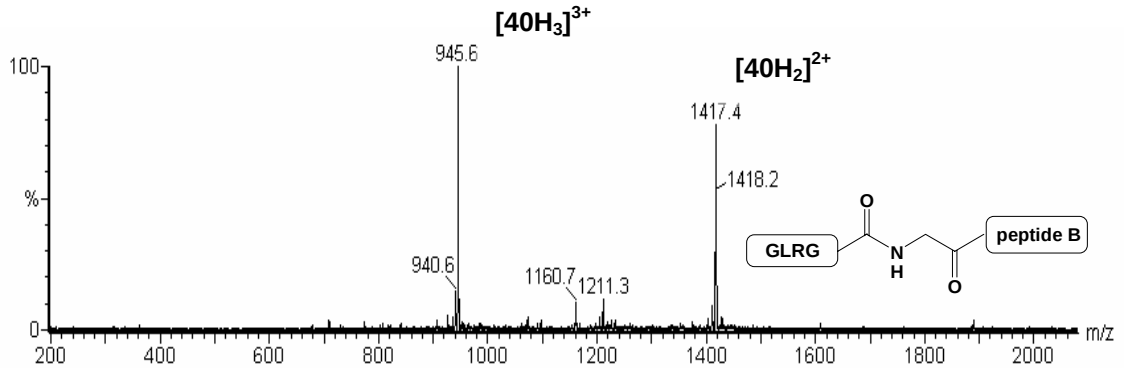


Figure 40: ESI positive ion mass spectrum of peptide **40** (calculated mass = 2833.2 Da, observed: 1417.4 Da $[MH_2]^{2+}$, 945.6 Da $[MH_3]^{3+}$). peptide B = KIIEGFVTGAEDIISGASRITKS.

When the same conditions were applied to the removal of the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4** from peptide **39a** however, a mixture of the desired cleaved product **40** and a species with the same mass as the starting peptide **39a** was formed. The peaks corresponding to this species were not reduced by prolonged exposure to TFA. Due to the formation of this TFA-stable species, the target peptide **40** was isolated in a lower yield of 58 % in this case.

Danishefsky and co-workers have since noted a similar effect[113] during TFA cleavage of these auxiliaries and have postulated that a reversal of the *S – N* acyl shift which occurs during ligation takes place to form thioester **41**, caused by irreversible protonation of the benzylamine. This problem could be avoided by methylation of the auxiliary thiol with

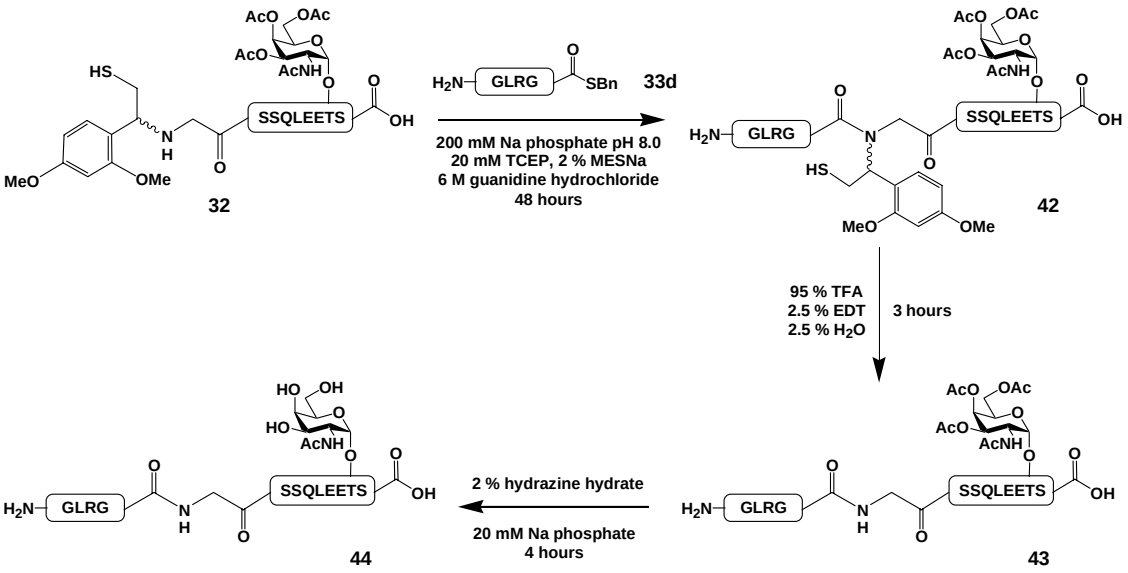
methyl *p*-nitrobenzene sulfonate prior to TFA treatment. As the TFA-stable byproduct was not isolated in our case however, it was not possible to confirm that thioester **41** had been formed.

Auxiliaries **4** and **6** were thus successfully used for traceless cysteine-free ligation at the Gly-Gly junction by ligation and subsequent TFA cleavage of the auxiliary. Auxiliary **6** was the more efficient of the two, both in ligation and in its removal, affording native peptide **40** in 71 % yield over the two steps, whereas removal of auxiliary **4** appears to have been complicated by the formation of a TFA-stable by-product.

3.1.2 Glycopeptide ligation

Following the successful ligations carried out for the Gly-Gly junction above, the more effective auxiliary **6** was used to assemble a glycopeptide fragment via ligation at the same junction in order to test the compatibility of the ligation and auxiliary removal conditions with glycosylated peptides.

Auxiliary-glycopeptide **32** was ligated to *C*-terminal glycine benzyl thioester **33d** under the same conditions as the for the previous Gly-Gly ligations, as shown in **Scheme 50**. The ligation was followed by LC-MS and was complete after 48 hours. The ligated glycopeptide product **42** was purified by HPLC as before.



Scheme 50: glycopeptide ligation and auxiliary removal.

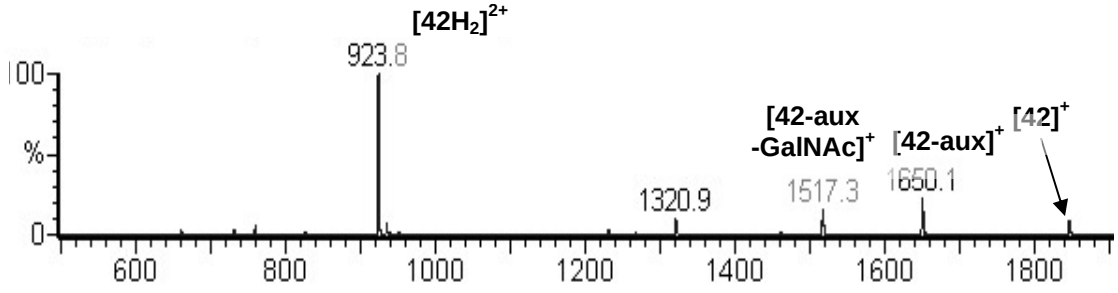


Figure 41: ESI positive ion mass spectrum from LC-MS analysis of glycopeptide ligation at time = 48 hours, showing glycopeptide product **42**, with characteristic signals due to fragmentation of auxiliary and the glycosidic linkage (calculated mass = 1845.9 Da, observed: 1845.9 Da $[M]^+$, 1650.1 Da $[M\text{-aux}]^+$, 1517.3 Da $[M\text{-aux-(OAc)}_3\text{GalNAc}]^+$, 923.8 Da $[MH_2]^{2+}$).

Ligation product **42** was treated with 95 % v/v TFA for three hours to remove the auxiliary exactly as before and the resulting crude product was purified by HPLC to afford the native peptide **43** (see **Figure 42**). **43** was then treated with 2 % v/v

hydrazine hydrate to remove the acetyl ester protecting groups from the saccharide to afford the fully deprotected native glycopeptide **44** as shown in **Figure 43**. This step also provided an additional confirmation of the formation of the stable amide-linked product, rather than the unrearranged thioester.

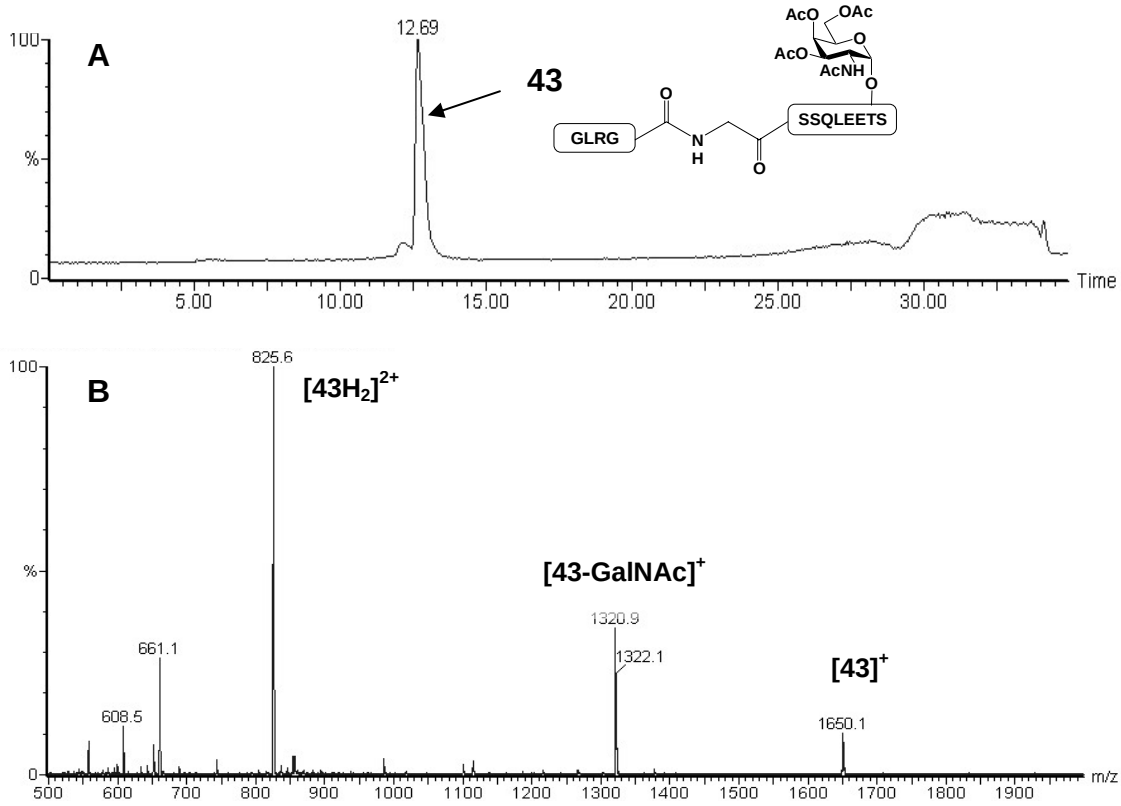


Figure 42: LC-MS analysis of HPLC purified glycopeptide **43**. **A**: TIC trace for purified **43** following auxiliary removal with TFA; **B**: ESI positive ion spectrum of **43**, again showing characteristic fragmentation of glycosidic linkage (calculated mass = 1649.7 Da, observed: 1650.1 Da $[M]^+$, 1320.9 Da $[M\text{-}(OAc)_3GalNAc]^+$, 825.6 Da $[MH_2]^{2+}$).

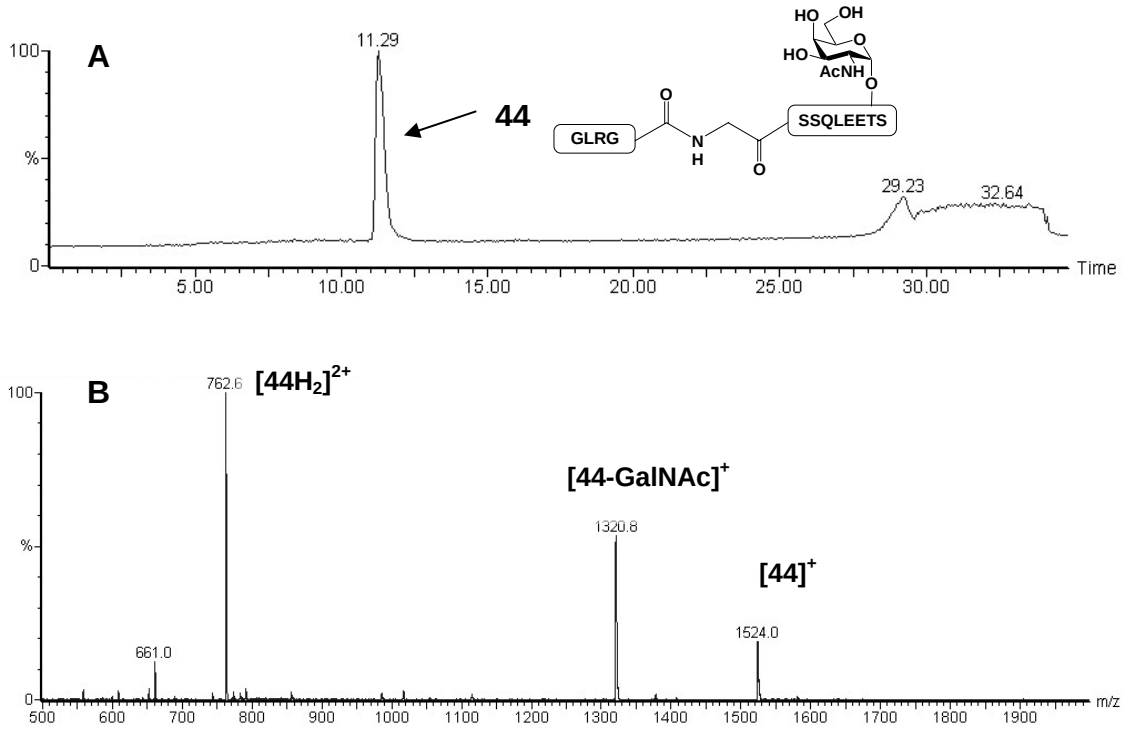


Figure 43: LC-MS analysis of HPLC purified deprotected glycopeptide **44**. **A**: TIC trace for purified **44** following acetyl ester removal with hydrazine; **B**: ESI positive ion mass spectrum of **44**, again showing characteristic fragmentation of glycosidic linkage (calculated mass = 1523.6 Da, observed: 1524.0 Da $[M]^+$, 1320.8 Da $[M-(OAc)_3GalNAc]^+$, 762.6 Da $[MH_2]^{2+}$).

No decomposition of the glycopeptide was observed during PMB deprotection of the auxiliary-glycopeptide or during the ligation, auxiliary removal and hydrazine deprotection steps, all of which proceeded cleanly and efficiently as monitored by LC-MS, thus confirming the compatibility of this auxiliary-mediated ligation approach with the assembly of glycopeptides.

3.2 Ligation at junctions other than Gly-Gly

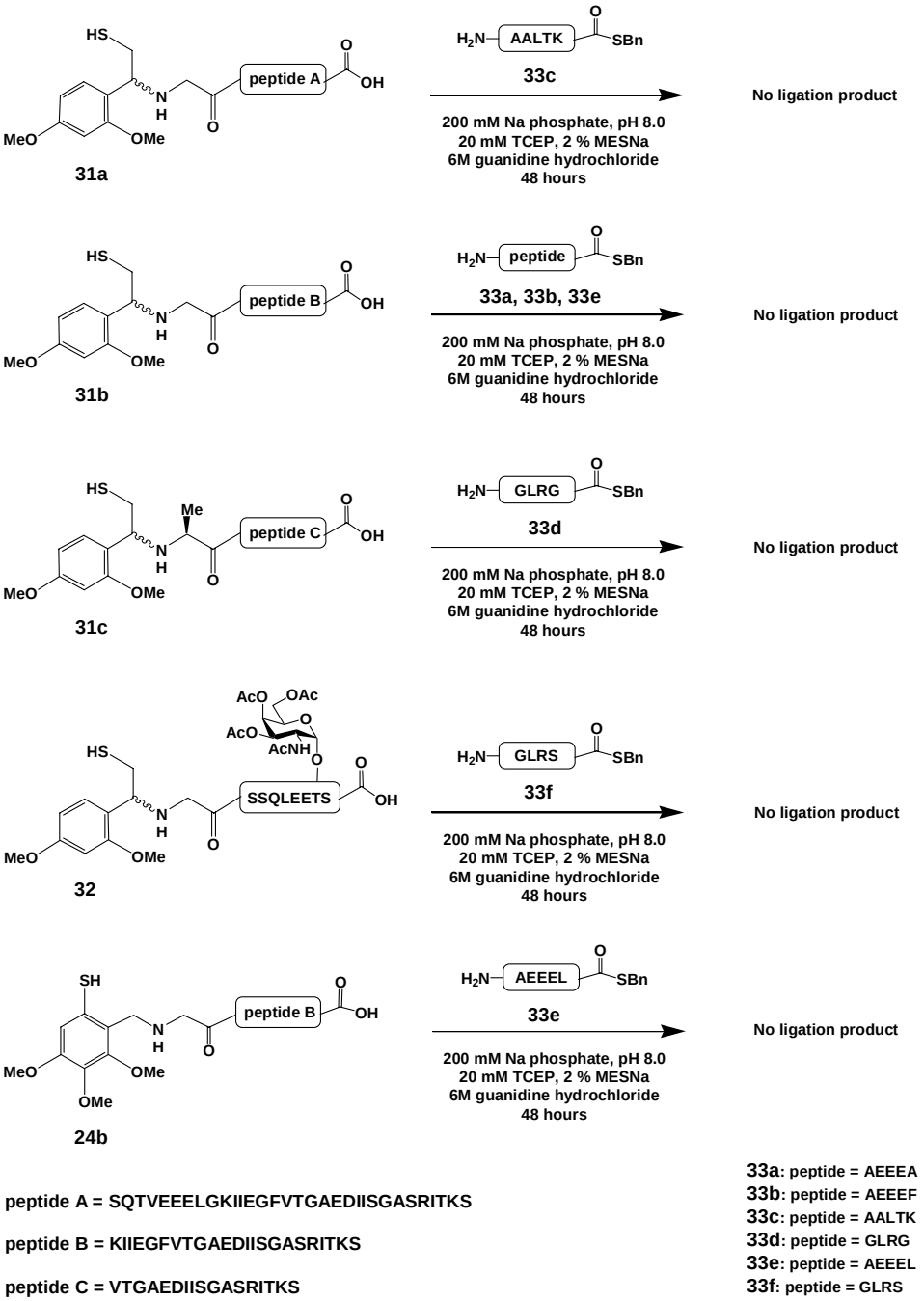
3.2.1 Initial studies

Following the success of the initial Gly-Gly ligation studies for the assembly of both peptides and glycopeptides, ligation studies were extended to a range of Xxx-Gly junctions. Peptides bearing the more effective *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** were thus ligated with short thioester peptides containing alanine and lysine as the *C*-terminal residue, as ligation at these junctions had been previously reported.[129,132,135] Studies were also conducted on the Ser-Gly, Leu-Gly and Phe-Gly junctions and the Gly-Ala junction. The first two of these were of particular interest as they are present at appropriate sites in the sequence of GlyCAM-1 for its assembly via auxiliary-mediated ligation with no changes in amino acid sequence.

Ligations were conducted under the same conditions as for the Gly-Gly ligations and monitored by LC-MS as before. Ligations of auxiliary-glycine peptides **31a** and **31b** with *C*-terminal alanine, phenylalanine, lysine and leucine benzyl thioesters **33a** – **33c** and **33e** were all unsuccessful, however (see **Scheme 51**).

Minor peaks corresponding to the mass of the ligated products were observed by LC-MS in the early stages of each ligation, but did not increase as the reaction progressed, therefore these signals were most likely due to the transthioesterified but unrearranged product. The findings in previous studies that the *S* – *N* acyl shift rearrangement is the

slowest and rate-determining step in such auxiliary-mediated ligations[128,129,132] support this assumption.



Scheme 51: unsuccessful Xxx-Gly and Gly-Ala ligations.

No further reaction was observed by LC-MS after 48 hours in each case. Ligation of auxiliary-glycopeptide **32** with *C*-terminal serine thioester **33f** was similarly unsuccessful. Ligation of the 4,5,6-trimethoxy-2-mercaptobenzyl-glycine peptide **24b** with *C*-terminal leucine thioester **33e** (see **Figure 44**) also failed, as did ligation at the Gly-Ala junction using *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)-alanine peptide **31c** and *C*-terminal glycine thioester **33d**.

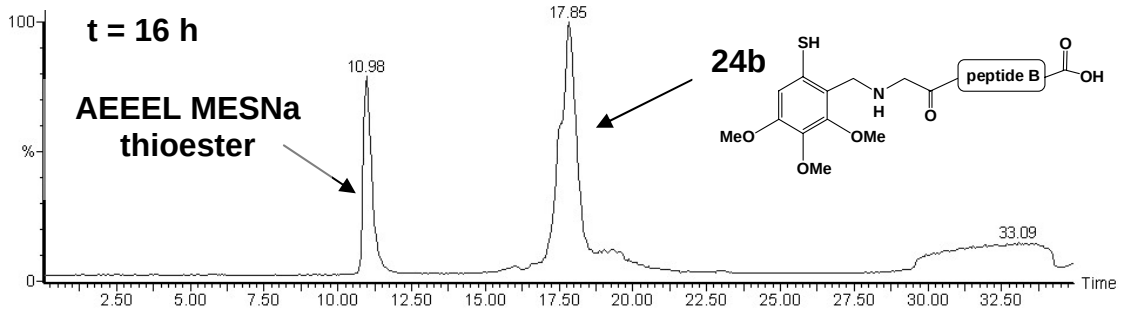


Figure 44: LC-MS TIC trace for ligation of auxiliary-peptide **24b** and *C*-terminal leucine thioester **33e** at time = 16 hours, showing only starting peptide **24b** and conversion of **33e** to MESNa alkyl thioester. peptide B = KIIEGFVTGAEDIISGASRITKS.

Although the steric bulk of the amino acid residues at the ligation junction is known to have a large effect on the speed of the critical $S - N$ acyl shift rearrangement in auxiliary-mediated ligations, the lack of any product formation for these non-Gly-Gly ligations was disappointing, especially for those junctions for which successful ligations had previously been reported. This was particularly true for the Ala-Gly junction, which is the least sterically-crowded non-Gly-Gly junction.

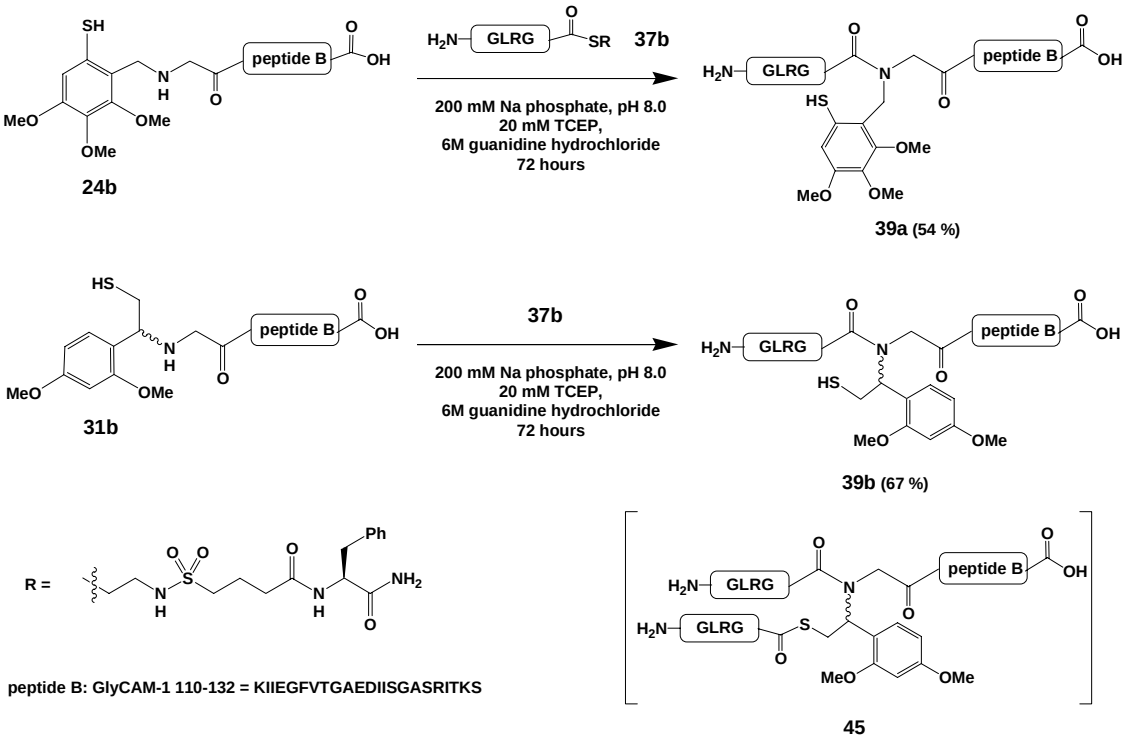
The original reports on the use of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6**[132,134] utilised very similar conditions to those used in our case: denaturing ligation buffer at basic pH, with reducing agent and thiol additives, although in this case the auxiliary was conjugated to only a short six-residue peptide. The authors of the paper describing the use of 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4**[129] however, noted an inhibitory effect of thiol additives on ligations, although no explanation for this effect was offered.

In each ligation reaction we had observed a rapid thiol exchange reaction between the benzyl or alkyl thioester and the MESNa catalyst to form the MESNa thioester. We therefore reasoned that the transthioesterified but unrearranged product would also be subject to similar rapid thiol exchange to regenerate the MESNa thioester and the starting auxiliary-peptide, and that the steric bulk around the non-Gly-Gly ligation junctions was slowing the critical rearrangement step to such an extent that the transthioesterified product was broken down before the rearrangement could occur. This underlines the dramatic effect that the steric bulk of the ligation junction has on the rearrangement step, rendering the ligation at the Ala-Gly junction ineffective in the presence of thiol, whereas no such problems were observed for the Gly-Gly junction.

3.2.2 Thiol-free ligations

A series of test ligations was conducted under the same conditions as previously described, but without the addition of MESNa. Following the observation of the

dramatic effect of steric bulk at the ligation junction, studies were concentrated on the less sterically-demanding Ala-Gly, Gly-Ala and Lys-Gly junctions. To determine the feasibility of ligation without the thiol catalyst, the Gly-Gly ligations with both auxiliaries were first repeated in the absence of MESNa.



Scheme 52: thiol-free ligations at the Gly-Gly junction.

Auxiliary-peptides **24b** and **31b** were thus ligated with alkyl thioester **37b** as before, but with no thiol additive (see **Scheme 52**). In both cases the ligations were more difficult to follow by LC-MS, as the spectra were frequently complicated by the presence of additional signals. Some of these corresponded to further thioester species, for example the double-ligated product **45** arising from acylation of the auxiliary thiol of the ligated product with a second molecule of the thioester (see **Figure 46**). Reaction progress was

also difficult to monitor due to the overlapping signals of the auxiliary-peptides and the ligated products in the LC-MS chromatograms.

To avoid possible undesired side reactions,[107] the reaction time was limited to 72 hours, after which time the reaction appeared to be finished as no further consumption of the thioester was observed. Two distinct peaks were again observed for the diastereomers of auxiliary-peptide **31b** and the resulting ligated product **39b**. This effect was also observed in subsequent ligations with *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6**.

To aid purification and break down any undesired thioester products, such as **45**, MESNa was added and the reaction mixture was shaken for a further six hours prior to HPLC purification. Ligated products **39a** and **39b** were isolated in 54 % and 67 % yield respectively. This corresponds to a slightly lower yield than for the ligations in the presence of MESNa catalyst for the entire duration of the reaction in both cases.

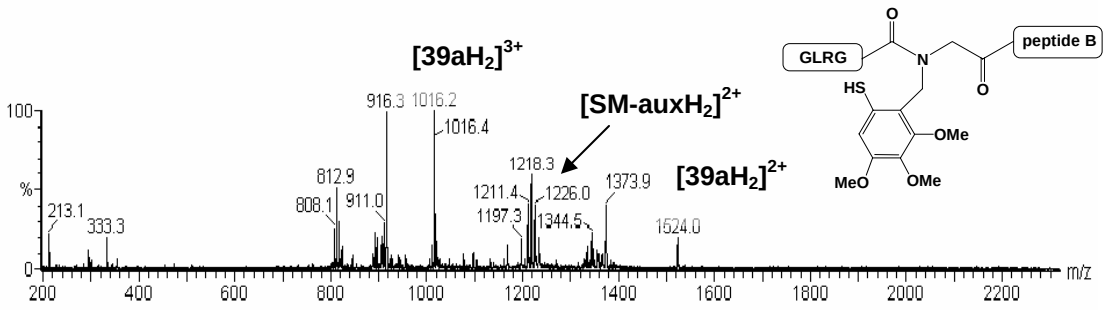


Figure 45: ESI positive ion mass spectrum of ligation product **39a** (calculated mass = 3045.5 Da, observed: 1524.0 Da $[MH_2]^{2+}$, 1016.2 Da $[MH_3]^{3+}$), showing trace of unreacted starting material. peptide B = KIIEGFVTGAEDIISGASRITKS.

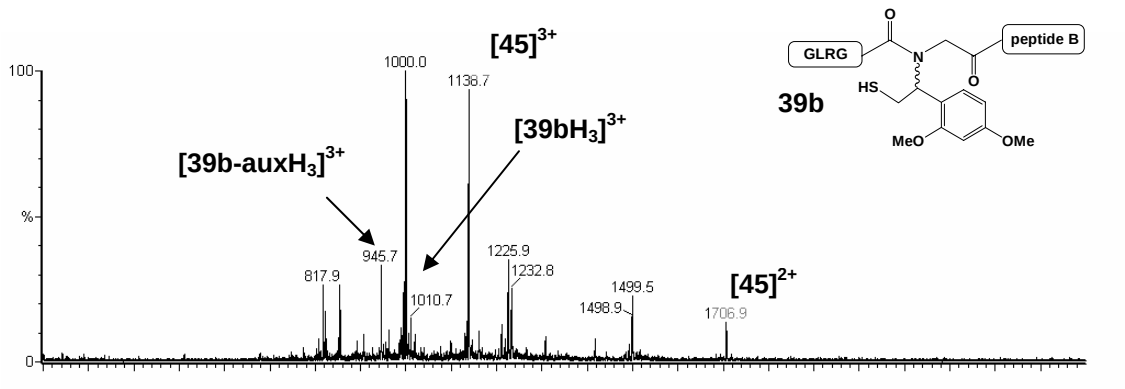
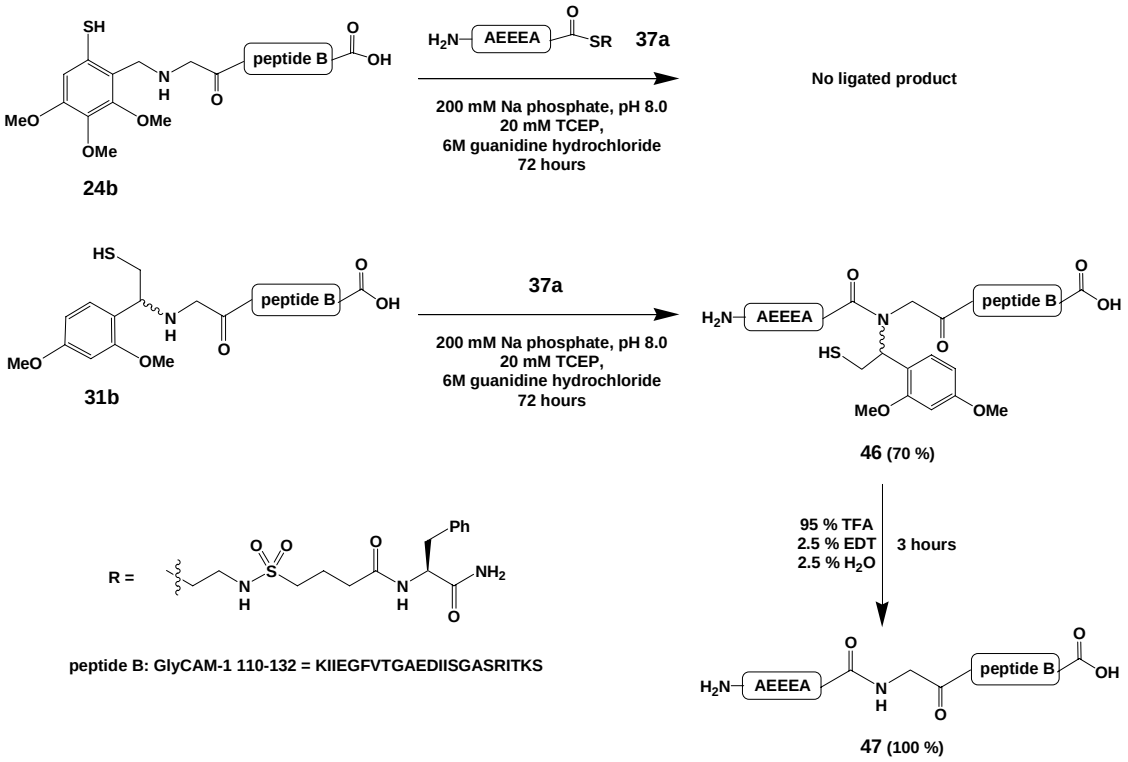


Figure 46: ESI positive ion mass spectrum of mixture of ligated product **39b** (calculated mass = 3029.5 Da, observed: 1010.7 Da $[MH_3]^{3+}$, 945.7 Da $[M\text{-}auxH_3]^{3+}$) and double-ligated product **45** (calculated mass = 3413.5 Da, observed: 1706.9 Da $[M]^{2+}$, 1138.7 Da $[MH_3]^{3+}$) from LC-MS analysis of ligation reaction mixture at time = 72 hours, prior to addition of MESNa. peptide B = KIIEGFVTGAEDIISGASRITKS.

The thiol free ligation conditions were then applied to the previously unsuccessful Ala-Gly ligations. Auxiliary-glycine peptides **24b** and **31b** were ligated to *C*-terminal alanine thioester **37a** for 72 hours in the absence of thiol, then for six hours with MESNa, and were followed by LC-MS as before (see **Scheme 53**).

Disappointingly, no product was observed after 72 hours ligation between **37a** and the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary peptide **24b**. Although thioester **37a** appeared to be consumed, no peaks corresponding to the ligated product were observed by LC-MS. Instead only peaks corresponding to the starting auxiliary-peptide were seen. *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** performed much better in this case, however. Ligated product **46** was isolated in 70 % yield by HPLC

from ligation between **31b** and **37a** for 72 hours, followed by addition of MESNa as before (see **Figure 47**). In contrast to the Gly-Gly ligation, the double-ligated product was not observed in this case.



Scheme 53: thiol-free Ala-Gly ligations.

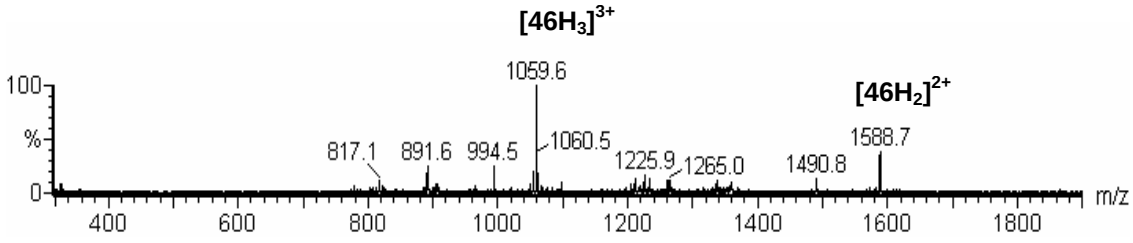


Figure 47: ESI positive ion mass spectrum of ligated product **46** (calculated mass = 3175.5 Da, observed: 1588.7 Da $[MH_2]^{2+}$, 1059.6 Da $[MH_3]^{3+}$).

Ligated product **46** was then treated with 95 % v/v TFA, 2.5 % v/v EDT to remove the auxiliary group (see **Scheme 53**). As seen for peptide **39b**, the reaction was complete within three hours as monitored by LC-MS and the native peptide product **47** was isolated in quantitiative yield by HPLC (**see Figure 48**).

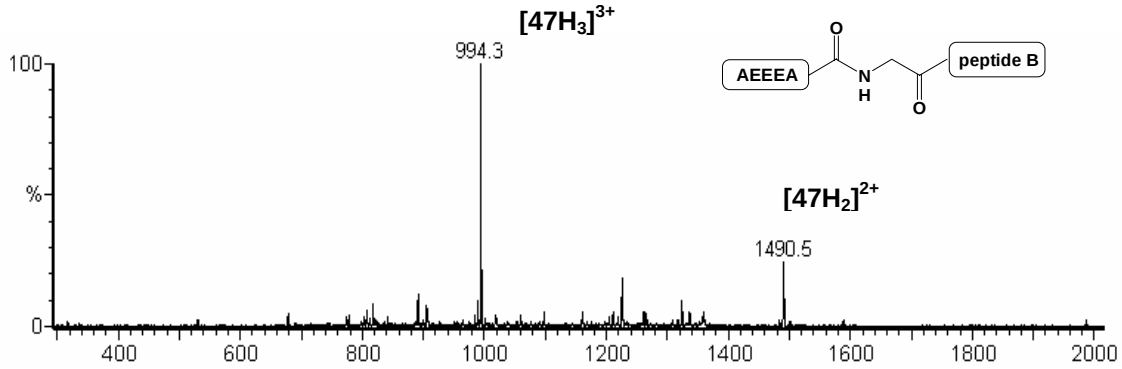
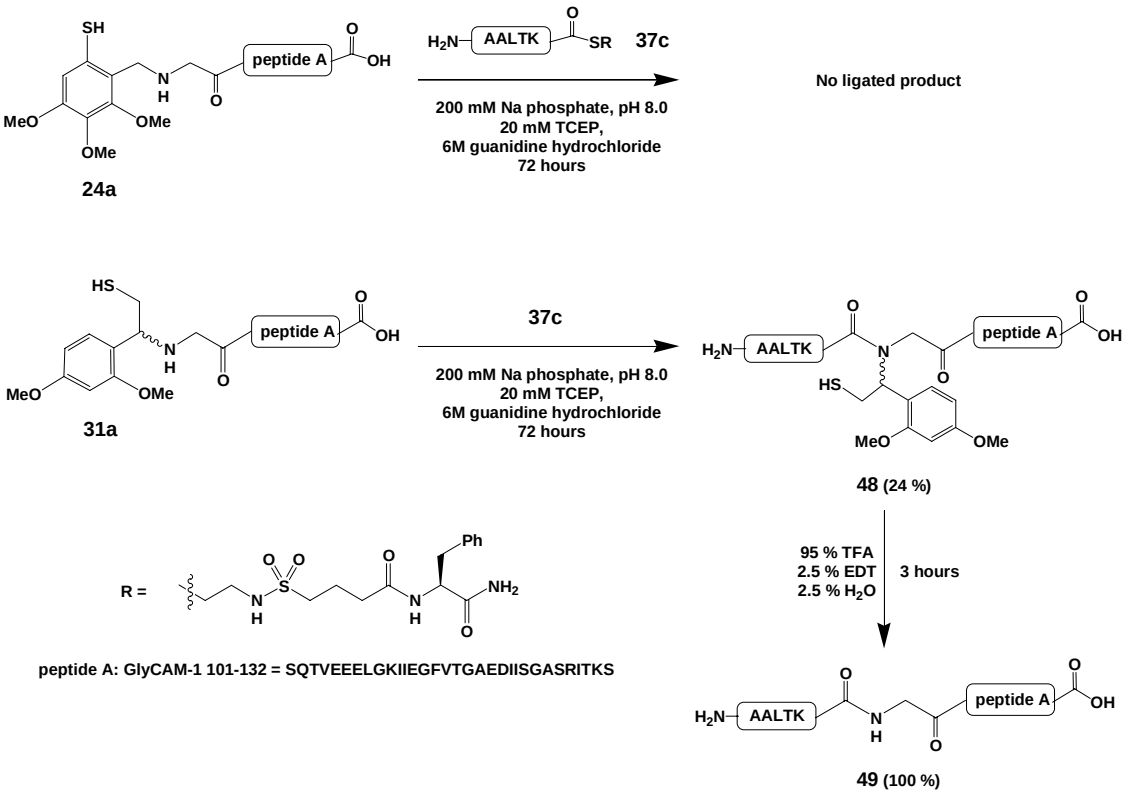


Figure 48: ESI positive ion mass spectrum of **47** (calculated mass = 2979.3 Da, observed: 1490.5 Da $[MH_2]^{2+}$, 994.3 Da $[MH_3]^{3+}$). peptide B = KIIEGFVTGAEDIISGASRITKS.

As for the Ala-Gly ligation, 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4** was unsuccessful at the Lys-Gly junction (see **Scheme 54**). No ligated product was observed by LC-MS after reaction of auxiliary peptide **24a** and *C*-terminal lysine thioester **37c** for 72 hours.

Ligation of *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary-peptide **31a** with **37c** for 72 hours however, resulted in a mixture of starting material **31a**, plus a smaller amount of ligated product **48**. Following treatement with MESNa, it was possible to isolate **48** by HPLC, although in a lower yield of 24 %. The auxiliary group was then removed by TFA cleavage as before to afford native peptide **49** in quantitative yield.



Scheme 54: thiol-free Lys-Gly ligations.

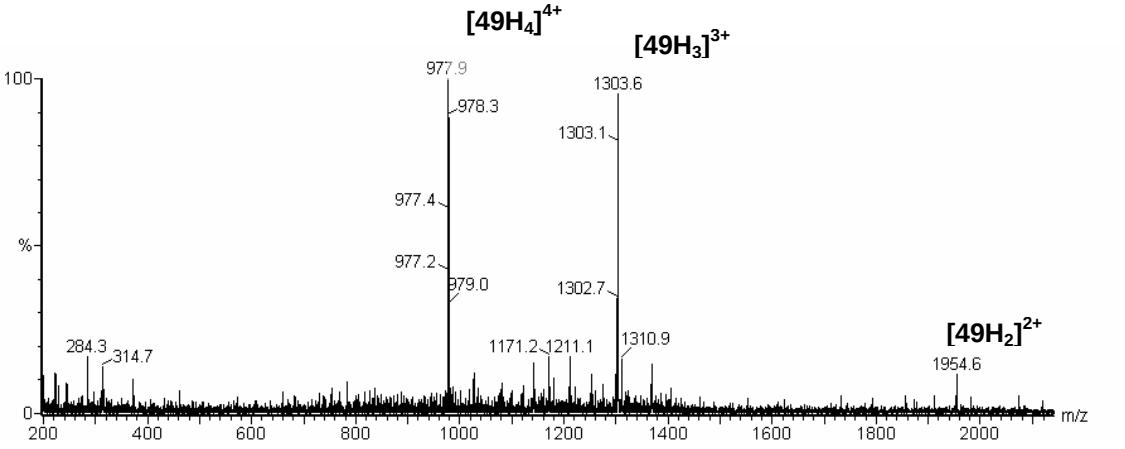
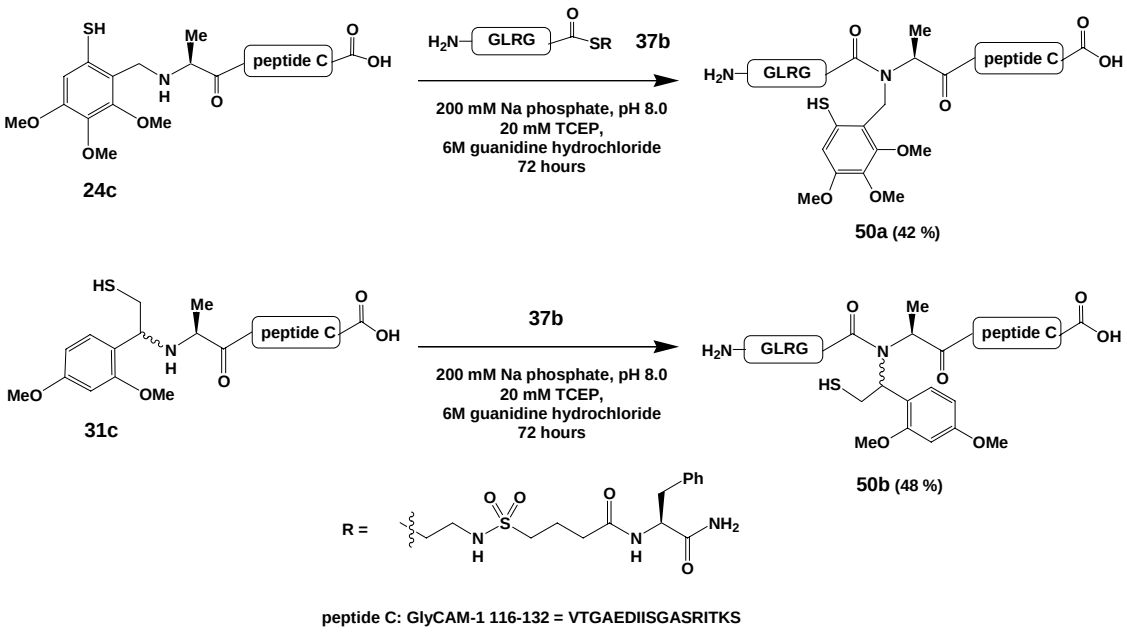


Figure 49: ESI positive ion mass spectrum of **49** (calculated mass = 3907.3 Da, observed: 1954.6 Da $[MH_2]^{2+}$, 1303.6 Da $[MH_3]^{3+}$, 977.9 Da $[MH_4]^{4+}$).

During the course of these studies, the assembly of peptides and cyclic peptides via ligation at a Gly-Ala junction using the HF-labile *N*α-(1-(4-methoxyphenyl)-2-mercaptoethyl) auxiliary **5** was reported.[133] Ligation at this junction had also been reported with 4,5-dimethoxy-2-mercaptobenzyl auxiliary **3**,[129] which also requires HF treatment for its removal. We were therefore interested in investigating whether the corresponding TFA-labile auxiliaries **6** and **4** would also be successful for ligations at this junction under the thiol-free conditions, as an attempted ligation at this junction with auxiliary **6** had failed in the presence of MESNa.



Scheme 55: thiol-free Gly-Ala ligations.

Auxiliary-alanine peptides **24c** and **31c** were thus ligated with glycine thioester **37b** for 72 hours, followed by MESNa addition for six hours, as before (see **Scheme 55**). Ligation was successful with both auxiliary-peptides, but in both cases analysis and purification were complicated by the presence of unreacted starting auxiliary-peptide,

which eluted very close to the ligated product. Products **50a** and **50b** were isolated in 42 % and 48 % yield respectively, containing traces of the corresponding starting auxiliary-peptide in both cases.

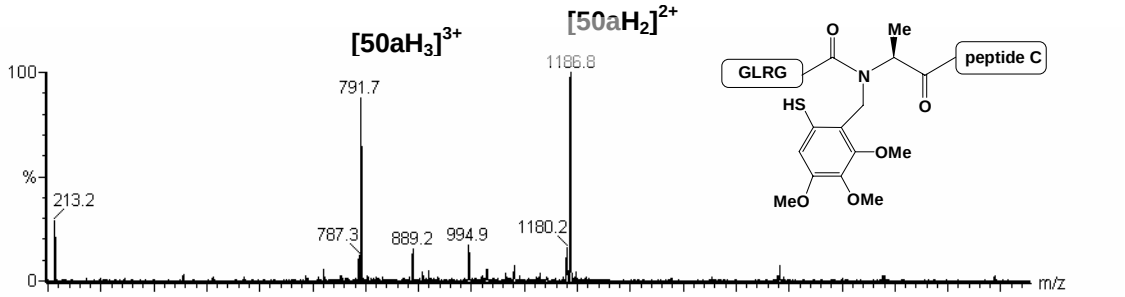


Figure 50: ESI positive ion mass spectrum of **50a** (calculated mass = 2371.7 Da, observed: 1186.8 Da $[MH_2]^{2+}$, 791.7 Da $[MH_3]^{3+}$).

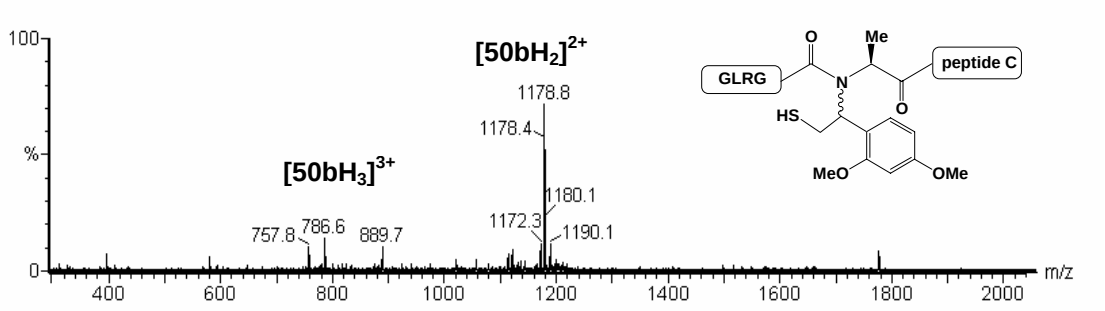
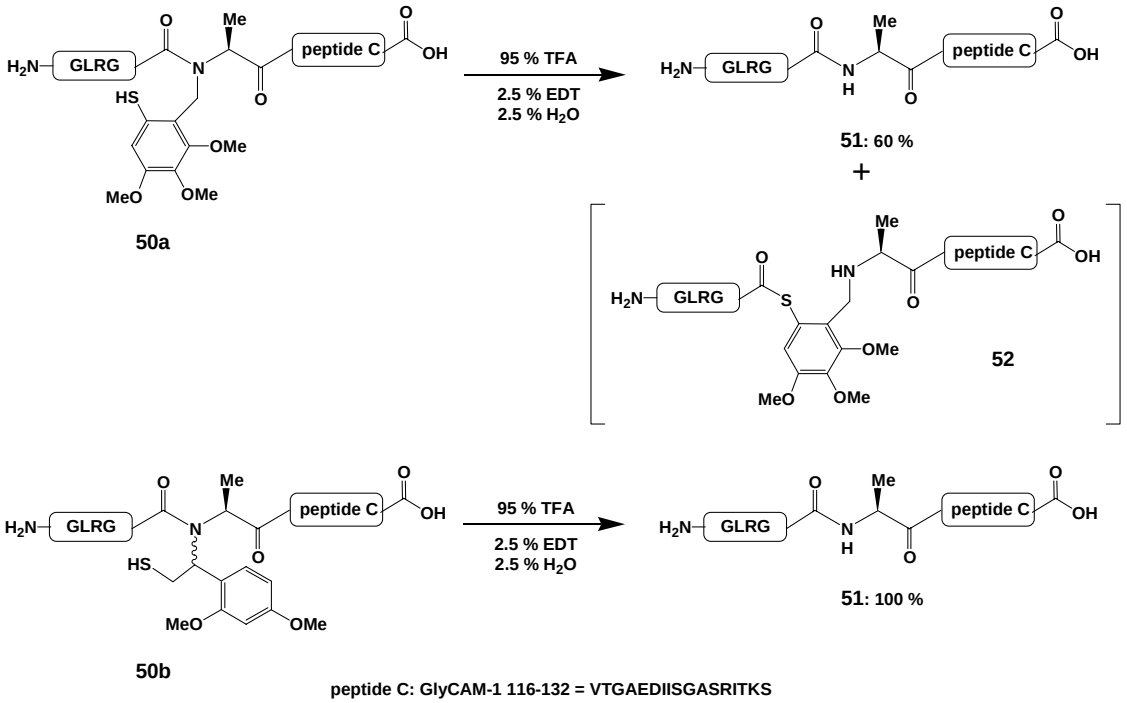


Figure 51: ESI positive ion mass spectrum of **50b** (calculated mass = 2355.7 Da, observed: 1178.8 Da $[MH_2]^{2+}$, 786.6 Da $[MH_3]^{3+}$). peptide C = VTGAEDIISGASRITKS.

Ligated products **50a** and **50b** were then treated with 95 % v/v TFA for three hours to remove the auxiliary groups (see **Scheme 56**). As previously seen, auxiliary **6** was removed from **50b** in quantitative yield by this procedure, whereas formation of an acid-

stable side product presumed to be the rearranged thioester **52** was observed during removal of auxiliary **4** from **50a**. Native peptide product **51** was therefore isolated in a lower yield of 60 % from TFA treatment of **50a**.



Scheme 56: auxiliary removal from ligated products **50a** and **50b**.

The auxiliary-peptide starting materials **24c** and **31c**, traces of which were present in the ligated products **50a** and **50b**, again proved difficult to remove during HPLC purification and traces of these starting materials were again detected in the native peptide product **51** (see **Figure 52**).

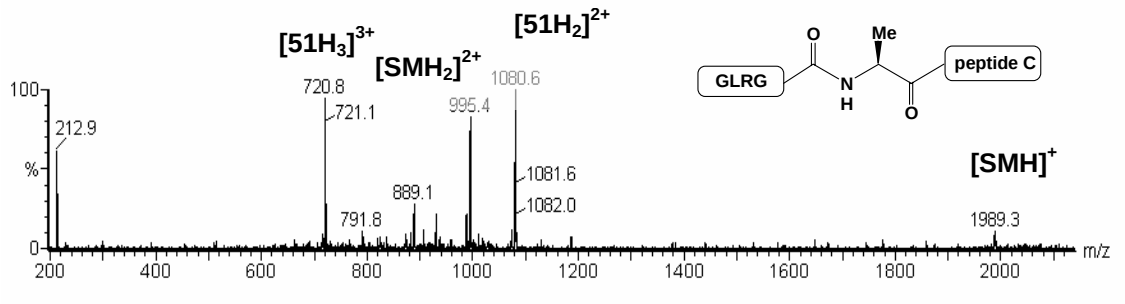
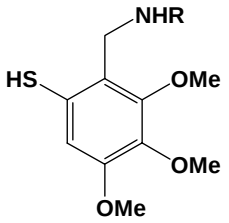


Figure 52: ESI positive ion mass spectra of peptide product **51** (calculated mass = 2159.4 Da, observed: 1080.6 $[MH_2]^{2+}$, 720.8 $[MH_3]^{3+}$) from LC-MS analysis of crude reaction mixture, showing traces of auxiliary-peptide starting material **4c**.

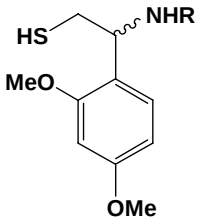
3.3 Ligation summary

Auxiliary-mediated ligations were thus successfully carried out at the Gly-Gly, Ala-Gly, Lys-Gly and Gly-Ala junctions under thiol-free conditions (see **Table 4**). The previously reported large effect of steric bulk at the ligation junction was observed for the non-Gly-Gly ligations. In these cases the rate-determining *S – N* acyl shift rearrangement was slowed to such an extent that ligation was unsuccessful in the presence of thiol additives, as the transthioesterified product was broken down by thiol exchange before rearrangement could occur.

For the most straightforward Gly-Gly ligation, better results were obtained in the presence of the MESNa thiol catalyst, as the formation of additional undesired thioester products occurred under thiol-free conditions.



4



6

Ligation junction	Auxiliary	Isolated yield (%)		Auxiliary Removal (%)
		+ MESNa	- MESNa	
Gly-Gly	**4**	64	54	58
Gly-Gly	**6**	71	67	100
Ala-Gly	**4**	0	0	n/a
Ala-Gly	**6**	0	70	100
Gly-Ala	**4**	0	42	60
Gly-Ala	**6**	0	48	100
Lys-Gly	**4**	0	0	n/a
Lys-Gly	**6**	0	24	100

Table 4: ligation summary.

Ligations using *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** were more successful for each ligation junction studied than those performed with the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4**. Ligation with auxiliary **6** at the Ala-Gly junction gave the best results obtained for thiol-free ligation, affording the ligated product in 70 % yield. An acceptable yield of 48 % was recorded for the converse Gly-Ala junction and it was also possible to achieve ligation at the Lys-Gly junction, albeit in a lower yield of 24 %. Auxiliary **6** could also be removed in quantitative yield in straightforward fashion from all of the ligation junctions studied, under conditions compatible with glycosylated peptides.

In contrast auxiliary **4** was unsuccessful in ligations at the Xxx-Gly junctions studied, although ligation at the Gly-Ala junction was successful in 42 % yield. Lower yields were also recorded for the removal of this auxiliary, due to the presumed rearrangement of the ligated product back to the TFA-stable thioester, as postulated by Danishefsky and co-workers.[113]

III. Conclusions

1. Conclusions

Rapid and facile routes to both the 4,5,6-trimethoxy-2-mercaptobenzyl **4** and *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) **6** TFA-cleavable auxiliaries have been developed, based on two main steps: firstly introduction of a protected thiol via either *S*-alkylation with the appropriate alkyl halide (for auxiliary **6**), or a palladium-catalysed *S*-arylation with aryl triflate (for auxiliary **4**), followed by direct reductive amination of the resulting aldehyde or ketone with an amino acid derviative.

PMB protected 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary-amino acid "cassettes" **11a**, **11c** and **11d** ready for use in SPPS were synthesised in four steps in good overall yields (54 % - 64 %) utilising a highly efficient microwave-assisted palladium catalysed aryl – thiol coupling reaction and a direct reductive amination with *t*-butyl protected amino acids. PMB and trityl protected *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary-amino acid cassettes **13a** – **13c** were synthesised in only two steps, again in good overall yields (53 % - 75 %) utilising direct reductive amination of initially unreactive ketones **12a** and **12b** with the unprotected amino acid under optimised conditions. Both of these syntheses represent considerable improvements over the previously published routes for such auxiliaries, being shorter and higher yielding and compatible with sensitive peptide modifications such as in glycoproteins, which are of particular interest to our group, due to the use of the PMB and trityl protecting groups.

The auxiliary-amino acid cassettes **11** and **13** were used directly in standard SPPS to efficiently introduce the auxiliaries to the *N*-termini of peptides, thus avoiding the extra steps and potentially problematic reactions involved in the previously employed "submonomer" conjugation approach. Fmoc protection of the auxiliary cassette was only necessary for 4,5,6-trimethoxy-2-mercaptobenzyl-glycine **11a**. It is noteworthy, however, that the auxiliary benzylamines **30a** and **30b** required for submonomer introduction of auxiliary **6** could also be formed in good yield by the above route, demonstrating the versatility of our approach. Following SPPS, the trityl and PMB protecting groups employed to mask the auxiliary thiol moeity were both efficiently removed under conditions compatible with glycosylated peptides, with trityl deprotection occurring concomitantly with release of the peptide from the resin. Very similar strategies for the assembly, protection and conjugation to peptides of these auxiliaries have since been used by other groups.

A "Double-linker" sulfonamide safety-catch strategy was applied for the study of the synthesis of the peptide thioester coupling partners. Mitsunobu alkylation, followed by intramolecular N to S acyl shift, was found to be the most effective method for the formation of the target thioester peptides, although repetetitive subjection of the activated sulfonamide resin to benzyl mercaptan cleavage may give comparable results.

Auxiliary-mediated ligations were successfully carried out at Gly-Gly ligation junctions in the presence of the MESNa thiol catalyst, to afford peptide and glycopeptide ligation products. The corresponding native peptide products were then obtained by straightforward removal of the auxiliary with TFA. Auxiliary **6** was removed in

quantitative yield by this procedure, but lower yields were recorded for the removal of auxiliary **4**. This effect may be due to the rearrangement of the ligated product back to the TFA-stable thioester, which has subsequently been proposed by Danishefsky and co-workers.[113] Removal of the acetyl ester protecting groups of the saccharide in glycopeptide product **43** was accomplished by treatment with hydrazine hydrate to afford fully deprotected native glycopeptide **44**. No decomposition of the glycopeptide was observed during the ligation, auxiliary removal and hydrazine deprotection steps, all of which proceeded cleanly and efficiently as monitored by LC-MS, thus confirming the compatibility of this auxiliary-mediated ligation approach with the assembly of glycopeptides.

Ligations at junctions other than Gly-Gly failed in the presence of MESNa. In these cases the rate-determining $S - N$ acyl shift rearrangement was slowed to such an extent by steric crowding at the ligation junction that ligation was unsuccessful in the presence of thiol additives, as the transthioesterified product was broken down by thiol exchange before rearrangement could occur.

Thiol-free ligation conditions were thus used for subsequent ligations and were successfully applied to ligations at the Gly-Gly, Ala-Gly, Lys-Gly and Gly-Ala junctions. For the most straightforward Gly-Gly ligation, better results were obtained in the presence of MESNa, as the formation of additional undesired thioester products occurred under thiol-free conditions.

Ligations using *N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) auxiliary **6** were more successful for each ligation junction studied than those performed with the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4**. Ligation with auxiliary **6** at the Ala-Gly junction gave the best results obtained for thiol-free ligation, affording the ligated product in 70 % yield. In contrast auxiliary **4** was unsuccessful in ligations at the Xxx-Gly junctions studied, although ligation at the Gly-Ala junction was successful in 42 % yield.

In summary, new, effective routes for the synthesis of auxiliaries **4** and **6** and their incorporation into peptides were developed and small-to-medium-sized peptides bearing the auxiliaries successfully participated in cysteine-free ligations at a range of ligation junctions. Introduction, deprotection and removal of the auxiliaries were also shown to be compatible with peptide glycosylation.

2. Future work

Future developments in this area will involve the application of sequential cysteine-free ligations to the assembly of large, complex glycopeptides and indeed impressive studies towards this goal have very recently been reported.[113] This report also details successful ligation at unfavourable Gly-Xxx junctions with the 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **4** under adapted ligation conditions utilising DMF cosolvent. The application of similar conditions seems likely to improve the efficiency of these ligations and extend the range of junctions at which ligation is possible. Model studies will be

required to determine which junctions are feasible for ligation with auxiliaries **4** and **6** under these conditions.

More efficient cleavage of auxiliary **4** was also described in the above report, via methylation of the auxiliary thiol to avoid the undesired reverse rearrangement to the thioester. The use of a reversible masking step, for example formation of the *S*-*tert*-butylthio- or 3-nitro-2-pyridinesulfenyl disulfides, would enable this procedure to be applied to cysteine-containing peptides.

The unique reactivity of the auxiliaries raises the possibility of their use in sequential ligation strategies without the need for protection of the thiol moiety. Existing sequential ligation strategies depend on temporary protection of *N*-terminal cysteine residues, which must be removed prior to subsequent ligation (see **Scheme 9**, page 44). As the auxiliaries are unreactive for non-Gly-Gly ligations in the presence of thiol additives, *N*-terminal auxiliary peptides could be assembled via a standard cysteine-ligation without the need for protection of the auxiliary thiol, which could then participate in a subsequent ligation under thiol-free conditions.

Very recent studies by Kent and co-workers have examined thioester reactivity in relation to rate of reaction for cysteine ligations.[107] Phenyl thioesters such as the 3-hydroxythiophenol thioester, similar to that formed by Danishefsky's *in situ* thioester formation approach, or the 4-(carboxymethyl)thiophenol thioester, were found to be more reactive than alkyl or benzyl thioesters. Although the rate-determining step for cysteine-free ligation is the $S - N$ acyl transfer rearrangement rather than the thiol

exchange step, faster formation of the transthioesterified product by use of these more reactive thioesters is likely to improve the efficiency of the ligation. Such thioesters are readily available via the use of the appropriate thiol in the cleavage step in the sulfonamide safety catch approach, or via a thiol exchange reaction with less reactive alkyl thioesters, such as those formed by the intramolecular thioester formation approach.

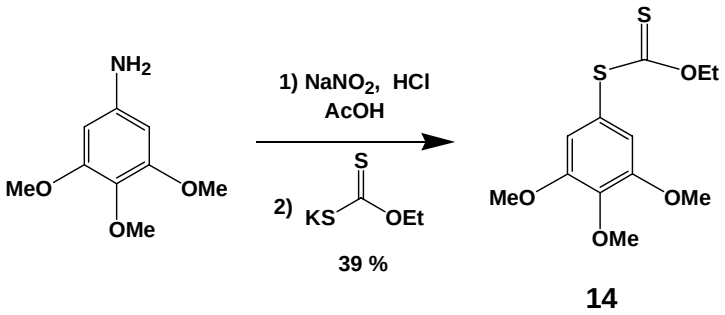
Experimental details

General

All reagents and solvents were purchased from Aldrich and Fisher and used as supplied unless otherwise stated. Anhydrous solvents were purchased from Aldrich and used as supplied. Analytical TLC was carried out on Merck aluminium-backed plates coated with silica gel. Plates were visualized using UV light and *p*-anisaldehyde or 2,4-dinitrophenylhydrazine dip. Flash chromatography was carried out on silica gel 60 Å particle size 35 – 70 micron.

^{1}H spectra were recorded at 250 MHz and ^{13}C NMR spectra at 63 MHz on a Bruker 250SY instrument. Mass spectra were obtained on a Micromass Platform II series electrospray MS and a Micromass Platform II mass spectrometer connected to a Waters Alliance HPLC system. LC-MS was performed using a Phenomenex Luna C18 LC-MS column (2.1×50 mm) and a gradient of 5-95 % acetonitrile containing 0.1% TFA over 25 minutes (flow rate of 0.2 ml/min). Semi-preparative HPLC was performed using a Phenomenex LUNA C18 column and a gradient of 10-90 % acetonitrile containing 0.1 % TFA over 50 minutes (flow rate of 3.0 ml/min).

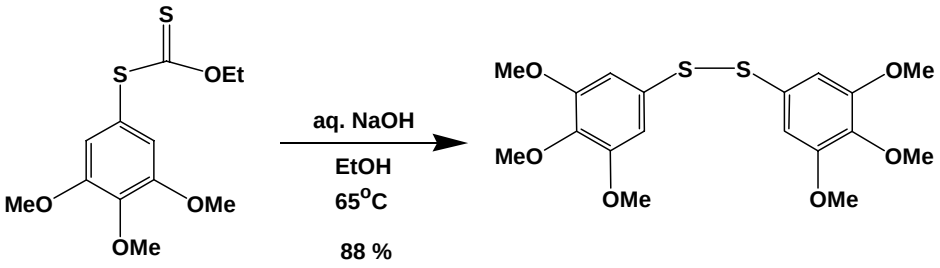
3,4,5-trimethoxyphenyl ethyl xanthate 14[129]



14

3,4,5-trimethoxyaniline (2.50 g, 13.7 mmol) was dissolved in acetic acid (8.0 ml). 10 % aqueous HCl (10.0 ml) was added and the reaction mixture was cooled to 0 ^{o}C. A solution of sodium nitrite (1.32 g, 19.2 mmol) in H_2O (5.0 ml) was added dropwise over 1 hour. Stirring was continued for a further 20 minutes at 0 ^{o}C. The reaction mixture was then added to a solution of potassium ethyl xanthate (4.38 g, 27.4 mmol) in H_2O (12.0 ml) at 65 ^{o}C. The reaction mixture was stirred at 65 ^{o}C for 20 minutes, then allowed to cool to room temperature. The reaction mixture was extracted with EtOAc (3 x 40.0 ml) and the combined organic extracts were washed with saturated NaCl solution (40.0 ml), dried over $MgSO_4$ and concentrated in vacuo. The crude residue was purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^{o}C, 10:90) to give the product **14** as a yellow oil (1.552 g, 5.38 mmol, 39 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 6.81 (2H, s, CH_{Ar} x 2), 4.69 (2H, q, J = 7.1 Hz, xanthate OCH_2), 3.95 (3H, s, OMe), 3.93 (6H, s, OMe x 2), 1.41 (3H, t, J = 7.1 Hz, xanthate CH_3). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 213.0 (qC, thiocarbonyl), 153.2 (qC x 2), 139.5 (qC), 122.1 (CH_{Ar} x 2), 70.1 (xanthate CH_2), 60.8 (CH_3, OMe), 56.2 (CH_3, OMe x 2), 13.5 (CH_3, xanthate CH_3). LRMS (ESI): found 310.8. $C_{12}H_{16}O_4S_2Na$ $[MNa]^+$ requires 311.04. ^{1}H and ^{13}C spectroscopic data in good agreement with literature (reference 129).

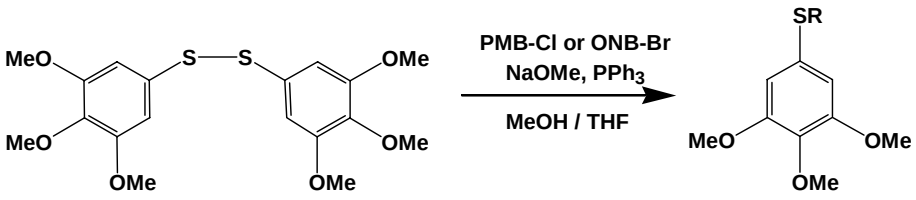
3,4,5-trimethoxy-thiophenol disulfide 15[129]



14 **15**

Xanthate **14** (1.5 g, 5.21 mmol) was dissolved in EtOH (20.0 ml). Aqueous NaOH solution (3 M, 20.0 ml) was added and the reaction mixture was stirred at 65 ^{o}C for 1 hour. The reaction mixture was allowed to cool, then acidified to pH 5 with 10 % v/v aqueous HCl and extracted with EtOAc (3 x 40.0 ml). The combined organic extracts were washed with saturated NaCl solution (40.0 ml), dried over $MgSO_4$ and concentrated in vacuo to afford **15** as a yellow oil (900 mg, 2.26 mmol, 88 % crude) which was used without further purification. LRMS (ESI): found 421.0. $C_{18}H_{22}O_6S_2Na$ $[MNa]^+$ requires 421.08. Not characterised in literature (reference 129).

3,4,5-trimethoxyphenyl thioethers 16a and 16b



15

16a: R = PMB (81 %)
16b: R = ONB (80 %)

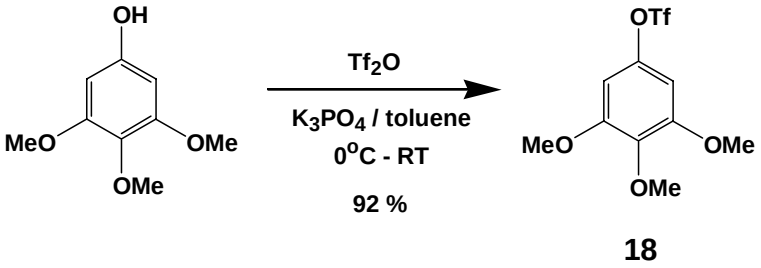
Disulfide **15** (860 mg, 2.16 mmol) was dissolved in a 1:1 v/v solution of THF/MeOH (50.0 ml). NaOMe (640 mg, 11.9 mmol) was added and the reaction mixture was stirred

for 5 minutes. Either *para*-methoxybenzyl chloride or *ortho*-nitrobenzy bromide (4.80 mmol) was added, followed by triphenylphosphine (1.16 g, 4.43 mmol) and stirring was continued for 15 minutes, after which the reaction mixture was concentrated in vacuo and purified by column chromatography over silica to give the products as white solids.

3,4,5-trimethoxyphenyl-(4-methoxybenzyl)-thioether 16a: column eluent EtOAc/petroleum ether bp 40-60 ^{o}C, 20:80 (1.117 g, 3.49 mmol, 81 %). Colourless solid, m.p. 64-65 ^{o}C. ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.07 (2H, d, J = 8.7 Hz, PMB CH_{Ar} x 2), 6.70 (2H, d, J = 8.7 Hz, PMB CH_{Ar} x 2), 6.40 (2H, s, Tmp CH_{Ar} x 2), 3.92 (2H, s, PMB CH_2), 3.70 (3H, s, OMe), 3.67 (3H, s, OMe), 3.66 (6H, s, OMe x 2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 158.6 (qC), 153.0 (qC x 2), 137.1, 130.5 (qC x 2), 129.9 (CH_{Ar} x 2), 129.6 (qC), 113.7 (CH_{Ar} x 2),, 108.3 (CH_{Ar} x 2), 60.7 (CH_3), 55.9 (CH_3 x 2), 55.1 (CH_3), 39.6 (CH_2). HRMS (FAB): found 320.1075. $C_{17}H_{20}O_4S$ $[M]^+$ requires 320.1082.

3,4,5-trimethoxyphenyl-(2-nitrobenzyl)-thioether 16b: column eluent EtOAc/petroleum ether bp 40-60 ^{o}C, 10:90 - 20:80 (1.155 g, 3.45 mmol, 80 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.84 (1H, dd, J_1 = 7.5 Hz, J_2 = 1.8 Hz, ONB CH_{Ar}), 7.28 (2H, td, J_1 = 7.5 Hz, J_2 = 1.8 Hz, ONB CH_{Ar} x 2), 7.07 (1H, dd, J_1 = 7.5 Hz, J_2 = 1.8 Hz, ONB CH_{Ar}), 6.32 (2H, s, Tmb CH_{Ar} x 2), 4.25 (2H, s, ONB CH_2), 3.67 (3H, s, OMe), 3.60 (6H, s, OMe x 2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 158.0 (qC x 2), 149.0, 138.9, 137.5, (qC x 3), 133.3, 131.4 (CH_{Ar} x 2), 131.3 (qC), 128.1, 124.9, 120.6, 105.3 (CH_{Ar} x 4), 62.3, 60.9, 55.9 (CH_3 x 3), 33.7 (CH_2). LRMS (ESI): found 336.4. $C_{16}H_{18}NO_5S$ $[MH]^+$ requires 336.09.

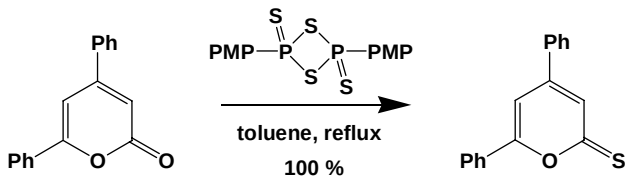
3,4,5-trimethoxyphenyl triflate 18



18

3,4,5-Trimethoxyphenol (2.00 g, 11.0 mmol) was dissolved in toluene (15.0 ml) and 30 % aqueous K_3PO_4 solution (15.0 ml) was added. The reaction mixture was cooled to 0 ^oC and trifluoromethansulfonic anhydride (2.20 ml, 13.0 mmol) was added slowly dropwise with stirring to maintain a reaction temperature of < 10 ^oC. The reaction mixture was allowed to warm to room temperature and stirred for a further 1 hour, then extracted with toluene (2 x 10.0 ml). The combined organic extracts were washed with water (30.0 ml), dried over $MgSO_4$ and concentrated in vacuo to give **18** as a pale yellow solid (3.23 g, 92 %) which was used without further purification. m.p. 62-63 ^oC. IR (nujol) ν_{max} 1420 cm^{-1}. ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 6.58 (2H, s, CH_{Ar} x 2), 3.95 (6H, s, OMe x 2), 3.93 (3H, s, OMe). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 153.6 (qC x 2), 145.1, 128.9, 128.1 (qC x 3), 99.9 (CH_{Ar} x 2), 60.9 (CH_3), 56.3 (CH_3 x 2). ^{19}F-NMR (235.3 MHz, $CDCl_3$): δ = -73.8. HRMS (FAB): found 316.0226. $C_{10}H_{11}F_3O_6S$ $[M]^+$ requires 316.0228.

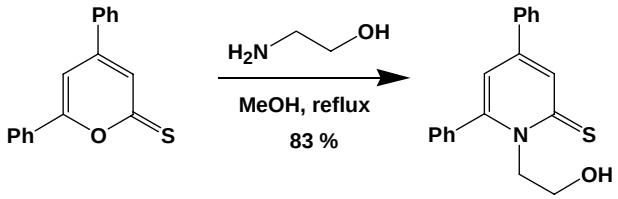
4,6-diphenylpyran-2-thione 21[185]



21

4,6-diphenyl-2-pyrone (1.00 g, 4.03 mmol) and Lawesson's reagent (3.26 g, 8.06 mmol) were suspended in anhydrous toluene (10.0 ml) under Ar. The reaction mixture was heated to reflux under a drying tube overnight. The reaction mixture was allowed to cool and then concentrated in vacuo and purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^{o}C, 20:80) to afford **21** as a red solid (1.06 g, 4.01 mmol, 100 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 8.00 (1H, m, CH_{Ar}), 7.71 (2H, m, CH_{Ar} x 2), 7.53-7.47 (7H, m, CH_{Ar} x 6 + CH), 7.20 (1H, d, J = 1.7 Hz, CH). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 196.8, 165.2, 148.3, 135.2 (qC x 4), 131.3 (CH), 130.6 (qC), 131.0, 129.3 (CH x 2), 129.0 (CH x 2), 126.8 (CH x 2), 125.9 (CH x 2), 125.4 (CH), 104.9 (CH x 2). LRMS (ESI): found 265.1. $C_{17}H_{13}OS$ $[MH]^+$ requires 265.07.

1-(2-hydroxyethyl)-4,6-diphenylpyridine-2-thione 20[185]

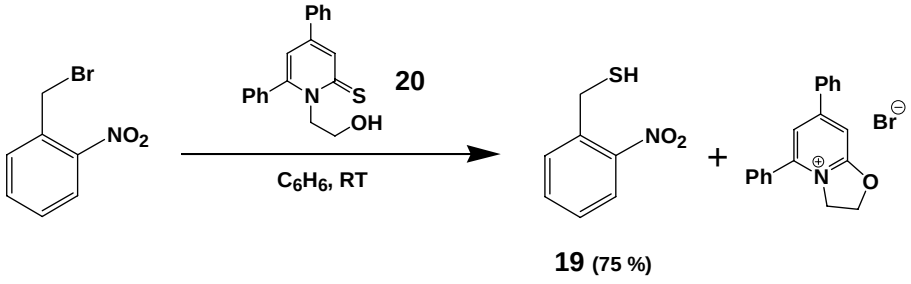


21 **20**

21 (1.05 g, 3.98 mmol) was dissolved in MeOH (60.0 ml). Ethanolamine (290 µl, 4.80 mmol) was added and the reaction mixture was heated to reflux for 7 hours. After this

time TLC analysis indicated incomplete reaction, so a further portion of ethanolamine (145 µl, 2.40 mmol) was added and heating was resumed for a further 5 hours. The reaction mixture was allowed to cool and concentrated in vacuo to give an orange solid, which was triturated with H_2O. The crude residue was then purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^oC, 30:70) to afford **20** as a yellow foam (871 mg, 2.84 mmol, 71 %). m.p 113-114 ^oC. IR (nujol) ν_{max} 3285, 1051 cm^{-1}. ^1H-NMR (250 MHz, $CDCl_3$): δ = 8.07 (1H, d, J = 2.2 Hz, CH), 7.62 (2H, m, CH_{Ar} x 2), 7.51 (3H, m, CH_{Ar} x 3), 7.45 (3H, m, CH_{Ar} x 3), 7.40 (2H, m, CH_{Ar} x 2), 6.38 (1H, d, J = 2.2 Hz, CH), 4.81 (2H, t, J = 5.5 Hz, CH_2), 3.91 (2H, t, J = 5.5 Hz, CH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 181.1, 153.0, 145.2, 135.7, 135.2 (qC x 5), 131.5, 130.0, 129.7 (CH x 3), 129.1 (CH x 2), 128.8 (CH x 2), 128.5 (CH x 2), 126.9 (CH), 115.6 (CH x 2), 61.7, 54.1 (CH_2 x 2). HRMS (FAB): found 308.1102. $C_{19}H_{18}NOS$ $[MH]^+$ requires 308.1109.

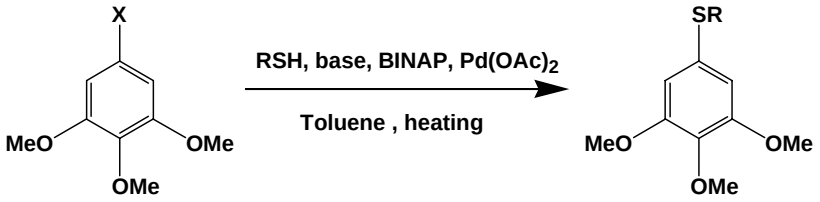
2-nitrobenzyl mercaptan 19[200]



20 (1.00 g, 3.26 mmol) was dissolved in anhydrous benzene (40.0 ml) and 2-nitrobenzyl bromide (700 mg, 3.24 mmol) was added. The reaction mixture was stirred under argon at room temperature for 24 hours, after which time a yellow precipitate of the pyridinium

bromide salt had formed. The precipitate was filtered off and washed with benzene (2 x 20.0 ml). The combined filtrates were concentrated in vacuo and purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^{o}C, 10:90 – 30:70) to afford **19** as a clear liquid (410 mg, 2.43 mmol, 75 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 8.02 (1H, m, CH_{Ar}), 7.57 (1H, m, CH_{Ar}), 7.45 (1H, m, CH_{Ar}), 7.35 (1H, m, CH_{Ar}), 4.02 (2H, d, J = 8.4 Hz, CH_2), 2.14 (1H, t, J = 8.3 Hz, SH). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 136.8 (qC), 132.9 (CH_{Ar}), 132.7 (qC), 131.4, 128.1, 125.4 (CH_{Ar} x 3), 26.3 (CH_2). LRMS (ESI): found 168.1. $C_7H_6NO_2S$ $[M-1]^-$ requires 168.01.

Typical procedure for microwave aryl-thiol coupling reactions (Table 2)



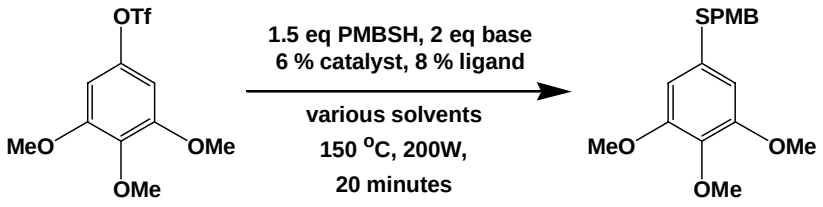
17: X = Br
18: X = OTf

16a: R = PMB
16b: R = ONB

200 mg of bromide **17** or triflate **18** was placed in a microwave tube and 11 mol % palladium diacetate and 11 mol % (*R*)-(+)-BINAP were added. The tube was sealed with a metal cap with septum, then toluene (5.0 ml or 2.5 ml in biphasic reactions), thiol (1.4 eq.) and base (1.4 eq. or 2.5 ml 10 % K_3PO_4) were added under argon. The tube was placed in a CEM Discover microwave operated with ChemDriver software and heated at 120 ^{o}C at a microwave power of 200 W for 20 minutes. The reaction mixture was diluted with ethyl acetate (25.0 ml) and washed with brine (15.0 ml), then dried over magnesium

sulfate, filtered and evaporated under vacuum. The crude products were purified by column chromatography over silica. **3,4,5-trimethoxy-(4-methoxybenzyl)-thioether 16a** was isolated in the yields detailed in **Table 2** by the above procedure, identical to **16a** produced by displacement of PMB chloride with disulfide **15** by TLC, NMR and MS. **3,4,5-trimethoxy-(2-nitrobenzyl)-thioether 16b** could not be isolated.

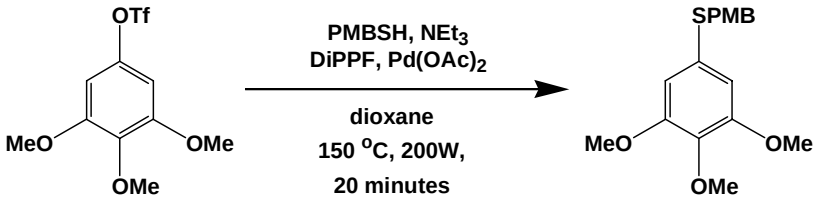
Preparation of samples for analytical HPLC (Table 3)



18 **16a**

Triflate **18** (50 mg, 0.16 mmol) was placed in a microwave tube and 6 mol % catalyst, 8 mol % ligand, base (0.32 mmol) and 4-methoxybenzyl mercaptan (35 µl, 0.24 mmol) were added. The tube was sealed with a metal cap with septum, then the solvent (2.5 ml) was added under argon. The tube was placed in a CEM Discover microwave operated with ChemDriver software and heated at 150 ^{o}C at a microwave power of 200 W for 20 minutes. 100 µl of the reaction mixture was diluted with MeOH (10.0 ml) and the resulting solution was filtered and analysed by HPLC against an standard sample of **16a**. Analytical HPLC was performed using a Phenomenex Luna C18 LC-MS column (250×5 mm) and a gradient of 10-90 % acetonitrile containing 0.1 % TFA over 45 minutes (flow rate of 0.1 ml/min) with UV detection at 254 nm.

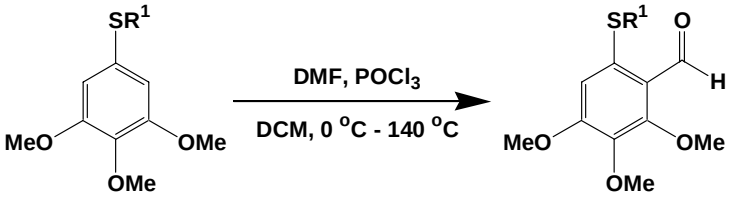
Optimised microwave aryl-thiol coupling



18 **16a** (97 %)

Triflate **18** (200 mg, 0.64 mmol) was placed in a microwave tube and DiPPF (20 mg, 0.05 mmol, 8 mol %), $Pd(OAc)_2$ (8 mg, 0.04 mmol, 6 mol %), NEt_3 (170 µl, 1.20 mmol) and 4-methoxybenzyl mercaptan (125 µl, 0.89 mmol) were added. The tube was sealed with a metal cap with septum, then anhydrous dioxane (5.0 ml) was added under argon. The tube was placed in a CEM Discover microwave operated with ChemDriver software and heated at 150 ^{o}C at a microwave power of 200 W for 20 minutes. The reaction mixture was diluted with ethyl acetate (25.0 ml) and washed with brine (15.0 ml), then dried over magnesium sulfate, filtered and evaporated under vacuum. The crude products were purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^{o}C, 20:80) to afford **16a** as a yellow solid (198 mg, 0.62 mmol, 97 %).

Vilsmeier formylation



16a: R^1 = PMB
16b: R^1 = ONB

10a: R^1 = PMB (90 %)
10b: R^1 = ONB (55 %)

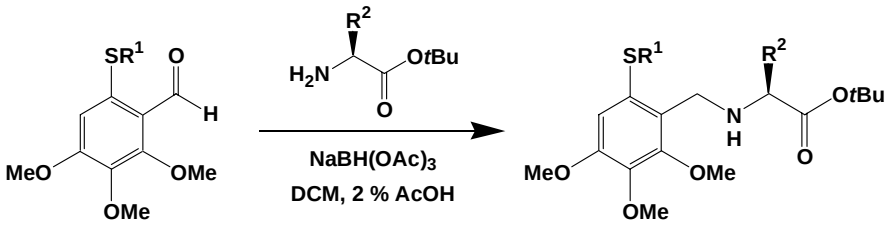
Thioether **16a** or **16b** (2.70 mmol) was dissolved in anhydrous DCM (5.0 ml) and added to a flame-dried flask fitted with a condenser under an argon atmosphere. Anhydrous DMF (0.21 ml, 2.70 mmol) was added and the reaction mixture was cooled to 0 ^{o}C. Phosphorous oxychloride (0.29 ml, 3.10 mmol) was added slowly dropwise over 10 minutes. The reaction was allowed to warm to room temperature, then heated to 150 ^{o}C for 5 hours. The reaction was allowed to cool, saturated sodium bicarbonate solution (20.0 ml) was added dropwise and the reaction was stirred at room temperature for 3 hours. The reaction mixture was then extracted with DCM (3 x 20.0 ml). The combined organic extracts were washed with brine (30.0 ml), dried over magnesium sulfate and evaporated under vacuum to give a brown solid, which was purified by column chromatography over silica to give the products **10a** or **10b**.

4,5,6-trimethoxy-2-(4-methoxybenzylthio)-benzaldehyde 10a: column eluent EtOAc/ petroleum ether bp 40-60 ^{o}C, 20:80 (white solid, 841 mg, 2.42 mmol, 90 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 10.26 (1H, s, aldehyde), 7.27 (2H, d, J = 8.6 Hz, PMB CH_{Ar} x 2), 6.77 (2H, d, J = 8.6 Hz, PMB CH_{Ar} x 2), 6.51 (1H, s, Tmb CH_{Ar}), 4.02 (2H, s, PMB CH_2), 3.90 (3H, s, OMe), 3.76 (3H, s, OMe), 3.74 (3H, s, OMe), 3.71 (3H, s, OMe). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 188.8 (qC, aldehyde), 158.8, 158.1, 157.9, 139.6, 138.1 (qC x 5), 129.7 (CH_{Ar}), 127.7, 119.9 (qC x 2), 113.9 (CH_{Ar} x 2), 104.4 (CH_{Ar} x 2), 62.2, 60.9, 55.9,

55.1 (CH_3 x 4, OMe x 4), 36.5 (CH_2). m/z (ESI): 371.0 $[MNa]^+$, $C_{18}H_{20}O_5SNa$ requires M 371.09.

4,5,6-trimethoxy-2-(2-nitrobenzylthio)-benzaldehyde 10b: column eluent EtOAc/ petroleum ether bp 40-60 ^oC, 20:80-30:70 (yellow solid, 539 mg, 1.48 mmol, 55 %). ^1H-NMR (250 MHz, $CDCl_3$): δ = 10.25 (1H, s, aldehyde), 7.75 (1H, d, J = 7.5 Hz, ONB CH_{Ar}), 7.61 (1H, d, J = 7.6 Hz, ONB CH_{Ar}), 7.47 (1H, t, J = 7.5 Hz, ONB CH_{Ar}), 7.37 (1H, t, J = 7.6 Hz, ONB CH_{Ar}), 6.38 (1H, s, Tmb CH_{Ar}), 4.42 (2H, s, ONB CH_2), 3.90 (3H, s, OMe), 3.56 (3H, s, OMe), 3.36 (3H, s, OMe). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 188.9 (qC, aldehyde), 158.0 (qC x 2), 149.0, 138.9, 137.5, (qC x 3), 133.3, 131.4 (CH_{Ar} x 2), 131.3 (qC), 128.1, 124.9 (CH_{Ar} x 2), 120.6 (qC), 105.3 (CH_{Ar}), 62.3, 60.9, 55.9 (CH_3 x 3), 33.7 (CH_2). LRMS (ESI): found 364.1. $C_{17}H_{18}NO_6S$ $[MH]^+$ requires 364.09.

Typical procedure for reductive amination of benzaldehydes



10a: R^1 = PMB (90 %)
10b: R^1 = ONB (55 %)

22a: R^1 = PMB, R^2 = H (93 %)
22b: R^1 = ONB, R^2 = H (93 %)
22c: R^1 = PMB, R^2 = Me (96 %)
22d: R^1 = PMB, R^2 = CH_2OtBu (94 %)

Benzaldehyde **10a** or **10b** (500 mg) was dissolved in DCM (40.0 ml) and acetic acid (0.80 ml, 2 %). The appropriate *t*-butyl protected amino acid (1.2 eq.) was added, followed by sodium triacetoxyborohydride (2.0 eq.). The reaction mixture was stirred at

room temperature for one hour, then neutralised with saturated sodium bicarbonate solution (50.0 ml). The reaction mixture was extracted with DCM (3 x 40.0 ml), washed with brine (50.0 ml), then dried over $MgSO_4$ and concentrated in vacuo to a yellow oil. The crude products were purified by column chromatography over silica to afford the pure products as white foams in yields of 93-96 %.

***N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzylglycine*t*-butyl ester 22a**: column eluent EtOAc/ petroleum ether bp 40-60 ^oC, 50:50 (93 %). ^1H-NMR (250 MHz, $CDCl_3$): δ = 7.08 (2H, d, J = 9.0 Hz, PMB CH_{Ar} x 2), 6.74 (2H, d, J = 9.0 Hz, PMB CH_{Ar} x 2), 6.58 (1H, s, Tmb, CH_{Ar}), 3.93 (2H, s, PMB CH_2), 3.83 (3H, s, OMe), 3.80 (3H, s, OMe), 3.78 (2H, s, Tmb CH_2), 3.73 (3H, s, OMe), 3.69 (3H, s, OMe), 3.21 (2H, s, Cα-H_2), 2.29 (1H, bs, NH), 1.39 (9H, s, ^{t}Bu). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 171.4, 158.6, 152.1, 152.3, 141.9, 130.3 (qC x 6), 129.9 (CH_{Ar} x 2), 129.2, 127.6 (qC x 2), 113.7 (CH_{Ar} x 2), 111.8 (CH_{Ar}), 80.7 (qC, ^{t}Bu), 61.1, 60.6, 55.8, 55.1 (CH_3 x 4, OMe x 4), 50.7, 44.8, 40.0 (CH_2 x 3), 27.9 (CH_3 x 3, ^{t}Bu). LRMS (ESI): found 463.8. $C_{24}H_{34}NO_6S$ $[MH]^+$ requires M 464.21.

***N*-4,5,6-trimethoxy-2-(2-nitrobenzylthio)benzylglycine*t*-butyl ester 22b**: column eluent EtOAc/ petroleum ether bp 40-60 ^oC, 40:60 (93 %). ^1H-NMR (250 MHz, $CDCl_3$): δ = 7.89 (1H, dd, J_1 = 7.5 Hz, J_2 = 1.8 Hz, ONB CH_{Ar}), 7.36 (2H, td, J_1 = 7.5 Hz, J_2 = 1.8 Hz, ONB CH_{Ar} x 2), 7.14 (1H, dd, J_1 = 7.5 Hz, J_2 = 1.8 Hz, ONB CH_{Ar}), 6.49 (1H, s, Tmb CH_{Ar}), 4.32 (2H, s, ONB CH_2), 3.84 (3H, s, OMe), 3.81 (3H, s, OMe), 3.76 (2H, s, Tmb, CH_2), 3.65 (3H, s, OMe), 3.22 (2H, s, Cα-H_2), 2.76 (1H, s, NH), 1.39 (9H, s, O*t*Bu). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 171.3, 152.6, 152.3 148.5, 142.5, 133.4 (qC x 6), 132.7, 132.2 (CH_{Ar} x 2), 129.1, 128.1 (qC x 2), 128.0, 124.9, 113.8 (CH_{Ar} x 3), 80.7 (qC, ^{t}Bu),

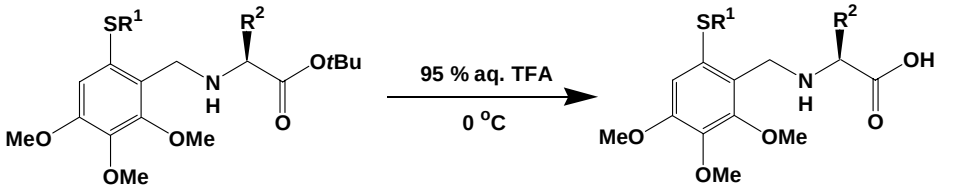
61.1, 60.6, 55.8 (CH_3 x 3, OMe x 3), 50.6, 45.1, 37.9 (CH_2 x 3), 27.9 (CH_3 x 3, ^{t}Bu). LRMS (ESI): found 479.2. $C_{23}H_{31}N_2O_7S$ $[MH]^+$ requires M 479.19.

***N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzyl-L-alanine*t*-butyl ester 22c**: column eluent EtOAc/ petroleum ether bp 40-60 ^{o}C, 50:50 (96 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 6.91 (2H, d, J = 8.6 Hz, PMB CH_{Ar} x 2), 6.69 (2H, d, J = 8.6 Hz, PMB CH_{Ar} x 2), 6.64 (1H, s, Tmb CH_{Ar}), 4.04 (2H, s, PMB CH_2), 3.89 (3H, s, OMe), 3.76 (4H, s + m, OMe + Cα-H), 3.69 (3H, s, OMe), 3.69 (3H, s, OMe), 3.66 (1H, d, J = 3.1 Hz, benzylic CH^a), 3.60 (1H, d, J = 3.1 Hz, benzylic CH^b), 1.47 (3H, d, J = 5.4 Hz, Cα-Me), 1.33 (9H, s, ^{t}Bu). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 167.9, 160.8, 157.4, 154.8, 141.5, 130.6 (qC x 6), 129.9 (CH_{Ar} x 2), 129.1, 118.2 (qC x 2), 113.7 (CH_{Ar} x 2), 113.4 (CH_{Ar}), 84.8 (qC, ^{t}Bu), 61.1, 60.7 (CH_3 x 2, OMe x 2), 55.9 (CH), 55.5, 55.0 (CH_3 x 2, OMe x 2), 44.1, 40.9 (CH_2 x 2), 29.5 (CH_3 x 3, ^{t}Bu), 15.1 (CH_3 Cα-Me). LRMS (ESI): found 478.1. $C_{25}H_{36}NO_6S$ $[MH]^+$ requires M 478.23.

***N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzyl-L-serine(*O*-*t*-butyl)*t*-butyl ester 22d**: column eluent EtOAc/ petroleum ether bp 40-60 ^{o}C, 30:70 (94 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.05 (2H, d, J = 9.0 Hz, PMB CH_{Ar} x 2), 6.71 (2H, d, J = 9.0 Hz, PMB CH_{Ar} x 2), 6.53 (1H, s, Tmb CH_{Ar}), 3.91 (2H, s, PMB CH_2), 3.81 (1H, d, J = 1.9 Hz, benzylic CH^a), 3.80 (3H, s, OMe), 3.77 (s, OMe), 3.74 (1H, d, J = 1.9 Hz, benzylic CH^b), 3.70 (3H, s, OMe), 3.64 (3H, s, OMe), 3.40 (2H, m, Cα-CH_2), 3.25 (1H, t, J = 5.9 Hz, Cα-H), 2.16 (1H, bs, NH), 1.38 (9H, s, *t*Bu), 1.06 (9H, s, *t*Bu). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 172.6, 158.6, 152.6, 152.1, 141.6 (qC x 5), 130.1 (CH_{Ar} x 2), 130.0, 129.5, 128.5 (qC x 3), 113.6 (CH_{Ar} x 2), 112.3 (CH_{Ar}), 80.4, 72.7 (qC x 2, ^{t}Bu x 2), 63.2 (CH_2), 61.9 (CH), 60.6, 60.2, 55.8, 55.1 (CH_3 x 4, OMe x 4), 44.3, 40.1 (CH_2 x 2), 28.0 (CH_3 x

3, ^{t}Bu), 27.2 (CH_3 x 3, ^{t}Bu). LRMS (ESI): found 551.1 $C_{29}H_{44}NO_7S$ $[MH]^+$ requires M 550.28.

Typical procedure for removal of *t*-butyl protecting groups



22a: R^1 = PMB, R^2 = H
22b: R^1 = ONB, R^2 = H
22c: R^1 = PMB, R^2 = Me
22d: R^1 = PMB, R^2 = CH_2OtBu

11c: R^1 = PMB, R^2 = Me (76 %)
11d: R^1 = PMB, R^2 = CH_2OH (71 %)

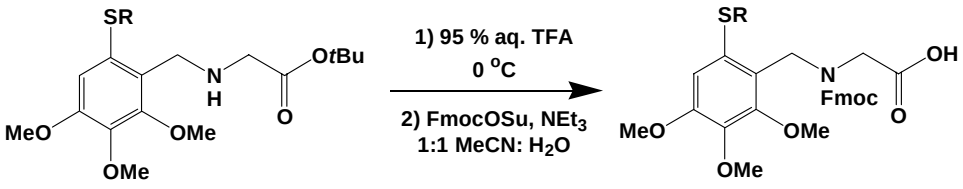
95 % TFA v/v , 5 % H_2O v/v (5.0 ml) was slowly added to the *t*-butyl protected auxiliary-amino acid (200 mg) at 0 °C. The reaction mixture was stirred for 3 - 4 hours at 0 °C until the reaction was complete by TLC analysis. The reaction mixture was then evaporated under vacuum and azeotroped with toluene. The crude auxiliary-alanine and serine acids **11c** and **11d** were then purified by column chromatography over silica.

***N*-4,5,6 - trimethoxy- 2-(4-methoxybenzylthio)benzyl alanine 11c**: column eluent EtOAc/ petroleum ether bp 40-60 °C, 70:30-100:0 (yellow foam, 76 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 8.38 (1H, bs, CO_2H), 6.96 (2H, d, J = 8.6 Hz, PMB CH_{Ar} x 2), 6.68 (2H, d, J = 8.5 Hz, PMB CH_{Ar} x 2), 6.49 (1H, s, Tmb CH_{Ar}), 4.14 (1H, d, J = 5.8 Hz, Tmb CH), 4.10 (1H, d, J = 5.8 Hz, Tmb CH), 3.85 (3H, s, OMe), 3.83 (2H, s, PMB CH_2), 3.72 (3H, s, OMe), 3.68 (3H, s, OMe), 3.62 (3H, s, OMe), 3.25 (1H, q, J = 7.0 Hz, Cα-H), 1.42 (3H, d, J = 7.0 Hz, Cα-Me). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 173.4, 158.8, 154.0, 152.6, 141.4, 130.8 (qC x 6), 130.0 (CH_{Ar} x 2), 129.0, 119.4 (qC x 2), 113.7 (CH_{Ar} x 2),

113.0 (CH_{Ar}), 61.2, 60.7, 55.8 (CH_3 x 3), 55.3 (CH), 55.1 (CH_3), 41.9, 40.5 (CH_2 x 2), 14.5 (CH_3). LRMS (ESI): found 421.9. $C_{21}H_{28}NO_6S$ $[MH]^+$ requires M 422.16.

***N*-4,5,6 - trimethoxy- 2-(4-methoxybenzylthio)benzyl serine 11d**: column eluent DCM/MeOH, 95:5-80:20 (white foam, 71 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 8.82-7.45 (1H, bs, CO_2H), 6.90 (2H, d, J = 8.5 Hz, PMB CH_{Ar} x 2), 6.68 (2H, d, J = 8.5 Hz, PMB CH_{Ar} x 2), 6.52 (1H, s, Tmb CH_{Ar}), 4.98 (1H, bs, NH), 4.08 (2H, AB quartet, Tmb CH x 2), 3.88-3.75 (5H, s, OMe + PMB CH_2), 3.71 (3H, s, OMe), 3.65 (3H, s, OMe), 3.61 (3H, s, OMe), 3.36 (3H, m, Cα-CH_2 + OH), 1.18 (1H, d, J = 8.0 Hz, Cα-H). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 172.0, 158.8, 154.4, 152.5, 141.4, 130.7 (qC x 6), 130.0 (CH_{Ar} x 2), 129.0, 118.6 (qC x 2), 113.7 (CH_{Ar} x 2), 113.4 (CH_{Ar}), 61.1 (CH), 60.7, 55.7, 55.0, 50.5 (CH_3 x 4), 43.0, 40.5, 29.6 (CH_2 x 3). LRMS (ESI): found 438.1. $C_{21}H_{28}NO_7S$ $[MH]^+$ requires M 438.16.

Typical procedure for Fmoc protection



22a: R = PMB
22b: R = ONB

11a: R = PMB (67 %)
11b: R = ONB (71 %)

The *t*-butyl protected auxiliary-glycine ester **22a** or **22b** was subjected to TFA deprotection exactly as described above for formation of **11c** and **11d**. Following azeotropic evaporation, the crude residue was dissolved in H_2O (5.0 ml). NEt_3 (1.0 eq. based on **22**) was added. 9-fluorenylmethoxycarbonyl-*N*-hydroxysuccinimide (1.0 eq.)

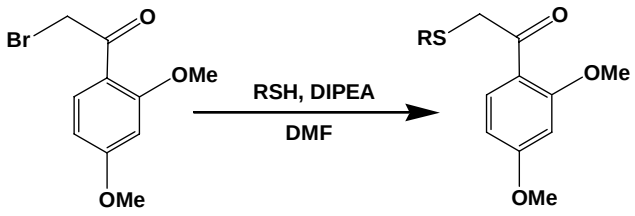
was dissolved in acetonitrile (5.0 ml) and added with stirring. The reaction mixture was stirred at room temperature for 3 hours, maintaining a pH of 8.5 - 9.0 by the addition of further NEt_3 as necessary. The reaction mixture was neutralized with 1.5 M HCl and extracted with DCM (3 x 20.0 ml). The combined organic extracts were washed with 1.5 M HCl (30.0 ml) and brine (30.0 ml), dried over $MgSO_4$, filtered and concentrated in vacuo to give the crude products as yellow oils, which were purified by column chromatography over silica to afford the products **11a** and **11b**.

***N*-9-fluorenylmethoxycarbonyl-*N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzyl-glycine 11a**: column eluent EtOAc/petroleum ether bp 40-60 ^oC, 90:10 (white foam, 67 %) m.p 72-74 ^oC. IR (nujol) ν_{max} 3018, 1718, 1699 cm^{-1}. ^1H-NMR (250 MHz, $CDCl_3$): δ = 7.54 (4H, m, Fmoc), 7.23 (4H, m, Fmoc), 7.00 (2H, d, J = 7.7 Hz, PMB CH_{Ar} x 2), 6.66 (2H, d, J = 7.7 Hz, PMB CH_{Ar} x 2), 6.53 (1H, d, J = 2.7 Hz, Tmb CH_{Ar}), 4.65 (2H, AB quartet, Tmb CH x 2), 4.39 (2H, d, J = 6.6 Hz, Fmoc CH_2), 4.18 (1H, m, Fmoc CH), 3.84 (2H, s, PMB CH_2), 3.73 (3H, s, OMe), 3.69 - 3.60 (11H, s x 4, OMe x 3 + CH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 174.7 (qC x2), 171.1 (qC), 158.7 (qC x2), 153.2 (qC x2), 143.9 (qC x2), 141.1 (qC x 4), 129.8 (CH_{Ar} x 2), 127.5 (CH_{Ar} x 2), 126.9 (CH_{Ar} x 2), 125.1, 124.8 (CH_{Ar} x 2), 119.9 (CH_{Ar} x 2), 113.7 (CH_{Ar} x 2), 68.1, 67.5 (CH_2 x 2), 60.6, 60.3, 55.8, 55.1 (CH_3 x 4), 60.3 (CH_2), 47.1 (CH), 43.1, 39.5 (CH_2 x 2). HRMS (FAB): found 629.2088. $C_{35}H_{35}NO_8S$ $[M]^+$ requires 629.2084.

***N*-9-fluorenylmethoxycarbonyl-*N*-4,5,6-trimethoxy-2-(2-nitrobenzylthio)benzyl-glycine 11b**: column eluent EtOAc/petroleum ether bp 40-60 ^oC, 90:10 (yellow gummy solid, 71 %). IR (nujol) ν_{max} 3018, 1716, 1697, 1526, 1352 cm^{-1}. ^1H-NMR (250 MHz, $CDCl_3$): δ = 7.80 (1H, dd, J_1 = 7.9 Hz, J_2 = 1.6 Hz, ONB CH_{Ar}), 7.66 (2H, m, Fmoc CH_{Ar} x 2), 7.51 (2H, m, Fmoc CH_{Ar} x 2), 7.36-7.11 (7H, m, Fmoc CH_{Ar} x 4 + ONB CH_{Ar} x 3),

6.46 (1H, s, Tmb CH_{Ar}), 4.58 (2H, AB quartet, Tmb CH x 2), 4.40 (2H, d, J = 5.6 Hz, Fmoc CH_2), 4.11-4.28 (3H, m + s, Fmoc CH + ONB CH_2), 3.75 (3H, s, OMe), 3.67 (3H, s, OMe), 3.61 (2H, s, CH_2), 3.63 (3H, s, OMe). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 174.3 (qC x 2), 171.1, 153.3, 148.5, 143.8 (qC x 4), 141.2 (qC x 4), 133.0 (qC), 132.9 (CH_{Ar} x 2), 132.1 (CH_{Ar}), 129.5 (qC), 128.1 (CH_{Ar} x 2), 127.5 (CH_{Ar} x 2), 126.9 (CH_{Ar} x 2), 125.1 (qC), 119.8 (CH_{Ar} x 2), 60.7 (CH_3 x 2), 60.3 (CH_2 x 2), 55.8 (CH_3), 47.1 (CH_2), 43.4 (Fmoc CH), 37.5 (CH_2 x 3). HRMS (FAB): found 667.1733. $C_{34}H_{32}N_2O_9SNa$ $[MNa]^+$ requires 667.1726.

Typical procedure for thiol displacement reactions



12a: R = PMB (98 %)
12b: R = Trt (96 %)

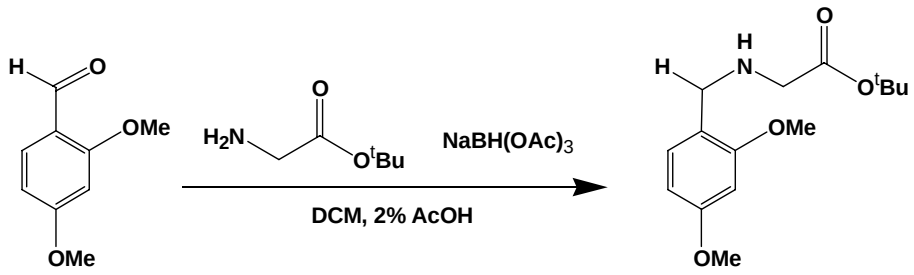
2-bromo-2,4-dimethoxyacetophenone (250 mg, 0.97 mmol) was dissolved in DMF (5.0 ml). Diisopropylethylamine (202 μl, 1.16 mmol) was added, followed by *para*-methoxybenzyl thiol or trityl thiol (1.16 mmol). The reaction mixture was stirred at room temperature for 2 hours, then evaporated to dryness in vacuo. The crude residue was dissolved in ethyl acetate (100 ml) and washed with brine (50.0 ml). The aqueous extract was extracted with ethyl acetate (50.0 ml) and the combined organic extracts were dried over $MgSO_4$, filtered and evaporated under vacuum to give yellow solids, which were

purified by column chromatography over silica to give the products **12a** and **12b** as white or yellow solids.

2-keto-2-(2,4-dimethoxyphenyl) (S-paramethoxybenzyl) ethanethiol 12a: column eluent EtOAc/petroleum ether bp 40-60 ^{o}C, 20:80 (white solid, 98 %). m.p. 69-71 ^{o}C. ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 8.03 (1H, d, J = 8.8 Hz, Dmb CH_{Ar}), 7.38 (2H, d, J = 8.0 Hz, PMB CH_{Ar} x 2), 6.95 (2H, d, J = 8.7 Hz, PMB CH_{Ar} x 2), 6.68 (1H, dd, J = 8.8 Hz, J = 2.3 Hz, Dmb CH_{Ar}), 6.58 (1H, d, J = 2.3 Hz, Dmb CH_{Ar}), 3.98 (6H, s x 2, OMe x 2), 3.91 (3H, s, OMe), 3.83 (2H, s, CH_2), 3.78 (2H, s, CH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 193.9, 164.7, 160.6, 158.5 (qC x 4), 133.3 (CH_{Ar}), 130.2 (CH_{Ar} x 2), 129.8, 119.0 (qC x 2), 113.7 (CH_{Ar} x 2), 105.4, 98.2 (CH x 3), 55.4 (CH_3 x 2), 55.1 (CH_3), 40.9, 35.0 (CH_3 x 2). HRMS (FAB): found 355.0973. $C_{18}H_{20}O_4SNa$ $[MNa]^+$ requires 355.0980.

2-keto-2-(2,4-dimethoxyphenyl) (S-trityl) ethanethiol 12b: column eluent EtOAc/petroleum ether bp 40-60 ^{o}C, 10:90-30:70 (pale yellow solid, 96 %). Decomposes at 121 ^{o}C. ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.60 (1H, d, J = 8.6 Hz, Dmb CH_{Ar}), 7.32 (6H, m, Trt), 7.12 (9H, m, Trt), 6.36 (1H, dd, J = 8.6 Hz, J = 2.3 Hz, Dmb CH_{Ar}), 6.19 (1H, d, J = 2.3 Hz, Dmb CH_{Ar}), 3.69 (3H, s, OMe), 3.51 (5H, s, OMe + CH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 193.9, 164.1, 160.1 (qC x 4), 144.1 (qC x 3), 132.5 (CH_{Ar}), 129.0 (CH_{Ar} x 6), 127.3 (CH_{Ar} x 6), 125.9 (CH_{Ar} x 3), 118.9 (qC), 104.6, 97.4 (CH_{Ar} x 2), 65.4 (qC), 54.9, 54.6 (CH_3 x 2), 44.6 (CH_2). HRMS (FAB): found 477.1503. $C_{29}H_{26}O_3SNa$ $[MNa]^+$ requires 477.1500.

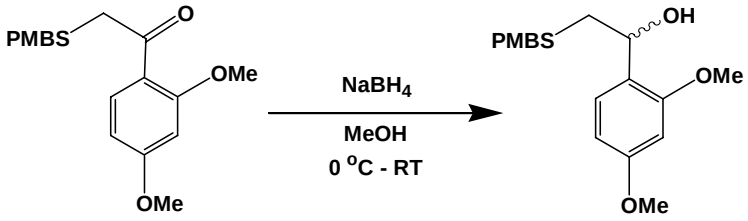
***N*-2,4-dimethoxybenzylglycine *t*-butyl ester 26**[198]



26 (94 %)

2,4-dimethoxybenzaldehyde (100 mg, 0.60 mmol) was dissolved in DCM (5.0 ml) containing acetic acid (100 µl, 2 %). *t*-butyl protected glycine (150 mg, 0.78 mmol) was added, followed by sodium triacetoxyborohydride (254 mg, 1.20 mmol). The reaction mixture was stirred at room temperature for 1 hour, then neutralised with saturated sodium bicarbonate solution (20.0 ml). The reaction mixture was extracted with DCM (3 x 20.0 ml), washed with brine (30.0 ml), then dried over $MgSO_4$, filtered and evaporated to a yellow oil. The crude products were purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^oC, 50:50-70:30) to afford **26** as a clear oil (159 mg, 0.57 mmol, 94 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.05 (1H, d, J = 8.1 Hz CH_{Ar}), 6.36 (1H, s, CH_{Ar}), 6.34 (1H, d, J = 2.4 Hz, CH_{Ar}), 3.79 (3H, s, OMe), 3.73 (3H, s, OMe), 3.66 (2H, s, Dmb CH_2), 3.21 (2H, s, Cα-H_2), 2.05 (1H, bs, NH), 1.38 (9H, s, *t*Bu). LRMS (ESI): found 282.0. $C_{15}H_{24}NO_4$ $[MH]^+$ requires 282.17. Spectroscopic data in good agreement with literature (reference 198).

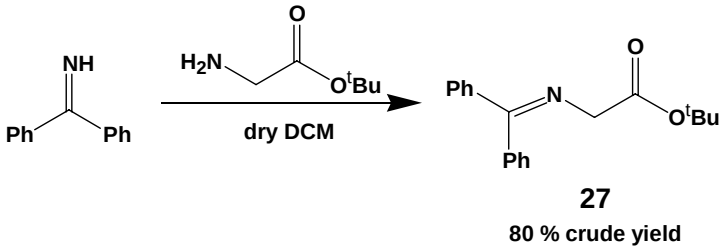
2-(4-methoxybenzyl)-2,4-dimethoxyacetophenyl alcohol 25



12a **25** (89 %)

Ketone **12a** (100 mg, 0.30 mmol) was dissolved in anhydrous MeOH and the reaction mixture was cooled to 0 ^{o}C. $NaBH_4$ (13 mg, 0.33 mmol) was added in 4 portions over 1 hour. The reaction mixture was stirred for a further 30 minutes at 0 ^{o}C, then allowed to warm to room temperature and stirring was continued for a further 1 hour. The reaction mixture was concentrated in vacuo, then partitioned between EtOAc (20.0 ml) and H_2O (20.0 ml) and stirred vigorously for 1 hour, then extracted with EtOAc (3 x 20.0 ml), dried over $MgSO_4$ and concentrated in vacuo. The crude product was purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^{o}C, 20:80) to afford **25** as a clear gummy solid (89 mg, 0.27 mmol, 89 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.45 (1H, d, J = 8.4 Hz, Dmb CH_{Ar}), 7.36 (2H, d, J = 6.6 Hz, PMB CH_{Ar} x 2), 6.98 (2H, d, J = 8.7 Hz, PMB CH_{Ar} x 2), 6.64 (1H, dd, J_1 = 8.4 Hz, J_2 = 2.4 Hz, Dmb CH_{Ar}), 6.57 (1H, d, J = 2.4 Hz, Dmb CH_{Ar}), 5.09 (1H, dd, J_1 = 8.6 Hz, J_2 = 4.0 Hz, Dmb CH), 3.94 (9H, s, OMe x 3), 3.80 (2H, s, PMB CH_2), 3.01 (1H, d x 2, J = 4.0 Hz), 2.76 (1H, d x 2, J = 8.6 Hz). LRMS (ESI): found 357.1128. $C_{18}H_{22}O_4SNa$ $[MNa]^+$ requires 357.1136.

***N*-diphenylmethyleneglycinetbutyl ester 27**[194]

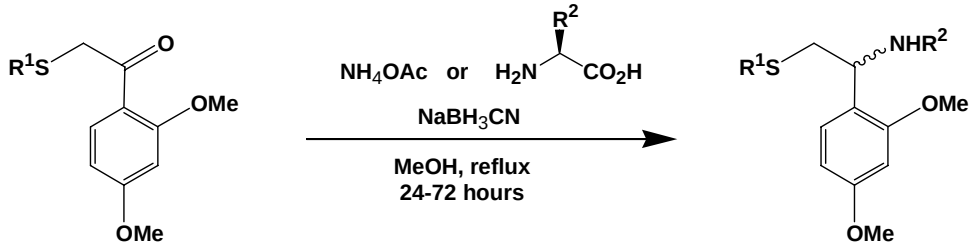


27

80 % crude yield

Glycine *t*-butyl ester (570 mg, 2.98 mmol) and benzophenone imine (0.50 ml, 2.98 ml) were dissolved in anhydrous DCM (10.0 ml) under a $CaCl_2$ drying tube and the reaction mixture was stirred at room temperature overnight. The reaction mixture was washed with brine (5.0 ml), dried over $MgSO_4$ and evaporated to dryness in vacuo to give the crude product **27** as a white solid (703 mg, 80 % crude yield) containing slight impurities by ^{1}H NMR and TLC. The crude product was therefore purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^{o}C, 0:100-10:90). **27** appeared to be unstable to chromatography however, as the starting imine was isolated as the major product. Almost all column fractions containing the desired product were also contaminated with the starting material. Chromatographic purification thus afforded only 17 mg (0.06 mmol, 2 %) of uncontaminated **27** as a white solid. m.p. 109-111 ^{o}C (literature 111-112 ^{o}C, reference 194). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.58 (2H, dd, J_1 = 7.9 Hz, J_2 = 1.3 Hz, CH_{Ar} x 2), 7.39 (3H, m, CH_{Ar} x 2), 7.26 (3H, m, CH_{Ar} x 3), 7.11 (2H, m, CH_{Ar} x 2), 4.04 (2H, s, Cα-H_2), 1.38 (9H, s, *t*Bu). LRMS (ESI): found 296.0. $C_{19}H_{22}NO_2$ $[MH]^+$ requires 296.17. Spectroscopic data in good agreement with literature (reference 194).

General procedure for reductive amination of acetophenones



12a: R^1 = PMB
12b: R^1 = Trt

13a: R^1 = PMB, R^2 = CH_2CO_2H (77 %)
13b: R^1 = Trt, R^2 = CH_2CO_2H (58 %)
13c: R^1 = Trt, R^2 = $CH(CH_3)CO_2H$ (55 %)
30a: R^1 = PMB, R^2 = H (79 %)
30b: R^1 = Trt, R^2 = H (61 %)

Ketone **12a** or **12b** (0.60 mmol) was dissolved in MeOH (5.0 ml). DCM (0.5 ml) was added to reactions with ketone **12b** to aid solubility. Ammonium acetate or glycine or alanine (6.00 mmol) was added, followed by sodium cyanoborohydride (27 mg, 0.42 mmol). The reaction mixtures were heated to reflux for 24 – 48 hours (for reaction with ammonium acetate) or 48 – 72 hours (for reaction with amino acid) then evaporated to dryness. With the exception of PMB benzylamine **30a**, the crude residues were purified by column chromatography over silica to give the products as white or yellow solids.

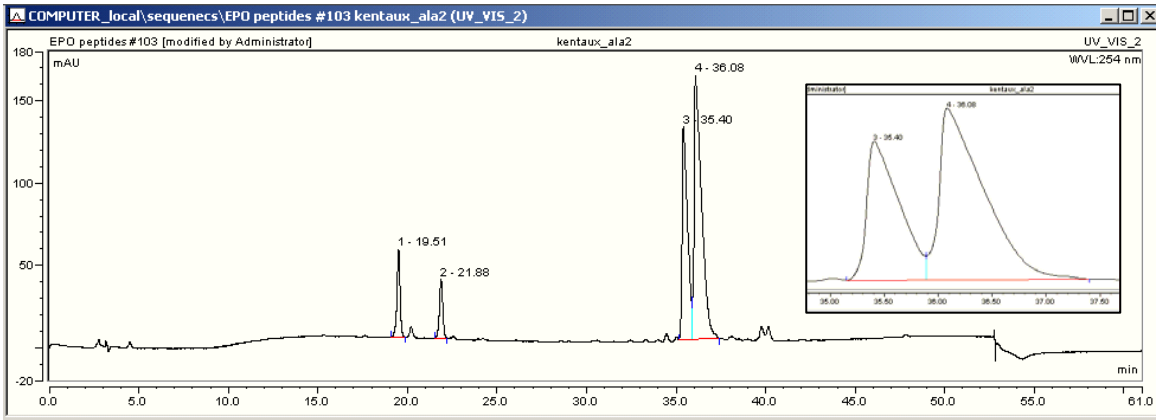
2-glycine-2-(2,4-dimethoxyphenyl) (S-paramethoxybenzyl) ethanethiol 13a: column eluent EtOAc/MeOH, 80:20-60:40 (pale yellow solid, 77 %) m.p 164-166 ^oC. IR (nujol) ν_{max} 3359, 3043, 1681 cm^{-1}. ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.11 (2H, d, J = 8.6 Hz, PMB CH_{Ar} x 2), 7.03 (1H, d, J = 8.5 Hz, DMB CH_{Ar}), 6.76 (2H, d, J = 8.6 Hz, PMB CH_{Ar} x 2), 6.43 (2H, m, DMB CH_{Ar} x 2), 4.13 (1H, m, DMB CH), 3.74 (9H, m, OMe x 3), 3.54 (2H, s, CH_2), 3.12 (2H, s, CH_2), 2.90 (2H, m, CH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 172.0, 165.2, 165.0, 161.8, 161.3 (qC x 5), 133.8 (CH_{Ar}), 133.0 (CH_{Ar} x 2), 132.0 (qC), 116.9 (CH_{Ar} x 2), 108.3, 102.0 (CH_{Ar} x 2), 61.1 (Dmb CH), 58.3 (CH_3 x 2) 58.0 (CH_3),

50.2, 38.1, 35.5 (CH_2 x 3). HRMS (FAB): found 392.1535. $C_{20}H_{26}NO_5S$ $[MH]^+$ requires 392.1532.

2-glycine-2-(2,4-dimethoxyphenyl) (S-trityl) ethanethiol 13b: column eluent EtOAc/MeOH, 100:0-60:40 (pale yellow solid, 58 %) m.p. 144-146 ^oC. IR (nujol) ν_{max} 3361, 3044, 1681 cm^{-1}. ^1H-NMR (250 MHz, $CDCl_3$): δ = 7.30-7.15 (5H, m, Trt), 7.17-7.06 (10H, m, Trt), 6.91 (1H, d, J = 8.4 Hz, Dmb CH_{Ar}), 6.35 (1H, dd, J_1 = 8.4 Hz, J_2 = 2.3 Hz, Dmb CH_{Ar}), 6.30 (1H, d, J = 2.3 Hz, Dmb CH_{Ar}), 3.51 (3H, s, OMe), 3.59 (3H, s, OMe), 3.45 (1H, m, Dmb CH), 3.04-2.88 (2H, m, CH_2), 2.86-2.76 (2H, m, CH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 171.1, 161.5, 158.1 (qC x 3), 144.2 (qC x 3), 131.3 (CH_{Ar}), 129.4 (CH_{Ar} x 6), 127.9 (CH_{Ar} x 6), 126.5 (CH_{Ar} x 3), 113.7 (qC), 104.4, 98.8 (CH_{Ar} x 2), 67.0 (qC), 60.3 (CH), 55.3 (CH_3 x 2), 47.3, 38.3 (CH_2 x 2). HRMS (FAB): found 514.2056. $C_{31}H_{32}NO_4S$ $[MH]^+$ requires 514.2052.

2-alanine-2-(2,4-dimethoxyphenyl) (S-trityl) ethanethiol 13c: column eluent EtOAc/MeOH, 90:10-70:30 (pale yellow solid, 55 %, 3:2 mixture of diastereomers) m.p. 153-156 ^oC. IR (nujol) ν_{max} 3360, 3041, 1685 cm^{-1}. ^1H-NMR (250 MHz, $CDCl_3$): δ = 7.55-7.40 (5H, m, Trt), 7.39-7.20 (10H, m, Trt), 7.00 (0.6H, d, J = 8.5 Hz, Dmb CH_{Ar}), 6.85 (0.4H, d, J = 8.5 Hz, Dmb CH_{Ar}), 6.54-6.43 (2H, m, Dmb CH_{Ar} x 2), 3.87 (1.8H, s, OMe), 3.85 (1.2H, s, OMe), 3.83 (1.2H, s, OMe), 3.74 (1.8H, s, OMe), 3.53 (1H, m, Dmb CH), 3.24-2.76 (3H, m, CH_2 + Cα-H), 1.29 (1.8H, d, J = 7.1 Hz, Cα-CH_3), 1.14 (1.2H, d, J = 7.1 Hz, Cα-CH_3). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 173.3, 162.0, 158.8, 144.9 (qC x 4), 113.1, 130.0, 128.4, 127.2, 127.0 (CH_{Ar} x 5), 114.1 (qC), 104.7, 99.6 (CH_{Ar} x 2), 67.5 (qC), 60.8 (CH), 55.8 (CH_3), 54.9 (CH), 34.3 (CH_2), 14.6 (CH_3). HRMS (FAB): found 528.2209. $C_{32}H_{34}NO_4S$ $[MH]^+$ requires 528.2208. HPLC analysis of diastereomeric mixture: 1.0 mg of the compound was dissolved in 1.0 ml of 20 % aqueous MeCN and 25

µl of the resulting solution was injected into an analytical Phenomenex C18 Luna reverse-phase column (4.6 mm x 250 mm). A gradient of 10 % - 85 % MeCN (containing 0.1 % v/v TFA) was applied over 40 mins (flow rate of 0.1 ml/min) with UV detection at 254 nm (see **Figure 53**). The two diastereomers eluted at Rt 35.40 and 36.08 min. with a peak area ratio of 43:57, which is in agreement with the diasteromeric ratio observed by NMR.



No.	Peakname	Ret.Time min	Area mAU*min	Amount	Type	Height mAU	Rel.Area %	Resolution
1	n.a.	19.506	10.7054	n.a.	BMB	53.045	7.60	7.58
2	n.a.	21.879	7.6587	n.a.	BMB*	35.715	5.44	30.14
3	n.a.	35.404	45.3252	n.a.	BM *	129.384	32.19	1.04
4	n.a.	36.084	77.1299	n.a.	MB*	160.337	54.77	n.a.
Total:			140.8193	0.0000		378.481	100.00	

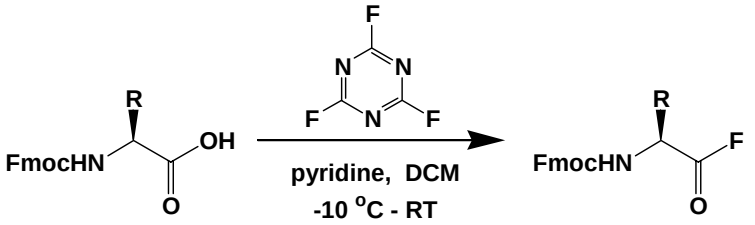
Figure 53: HPLC analysis of diastereomeric mixture for **13c**.

2-amino-2-(2,4-dimethoxyphenyl) (S-trityl) ethanethiol 30b: : column eluent EtOAc/MeOH, 100:0-70:30 (colourless foam, 61 %). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.12 (5H, m, Trt), 7.19 (10H, m, Trt), 6.78 (1H, d, J = 9.0 Hz, Dmb CH_{Ar}), 6.30 (2H, m, Dmb CH_{Ar} x 2), 3.69 (3H, s, OMe), 3.64 (3H, s, OMe), 3.55 (1H, t, J = 7.5 Hz, Dmb CH), 2.74 (2H, m, CH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 176.4, 160.9, 157.9 (qC x 3), 144.3, 129.8 (CH_{Ar}), 129.4 (CH_{Ar} x 6), 128.6 (CH_{Ar} x 6), 126.6 (CH_{Ar} x 3), 104.0 (CH_{Ar})

98.7 (CH_{Ar}), 67.0 (qC), 55.2, 55.0 (CH_3 x 2), 51.6 (CH), 35.2 (CH_2). LRMS (ESI): found 455.9. $C_{29}H_{30}NO_2S$ $[MH]^+$ requires 456.20.

2-amino-2-(2,4-dimethoxyphenyl) (S-paramethoxybenzyl) ethanethiol 30a: upon cooling the reaction mixture was acidified to pH ≈ 2 with concentrated HCl and washed with Et_2O (2 x 5.0 ml) to remove unreacted starting material. The aqueous fraction was then basified to pH ≈ 10 with powdered KOH and extracted with DCM (3 x 10.0 ml). The combined DCM extracts were dried over $MgSO_4$ and evaporated to dryness in vacuo to afford **30a** as a colourless foam (79 %) requiring no further purification. IR (nujol) ν_{max} 3368 cm^{-1}. ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.42 (1H, s, Dmb CH_{Ar}), 7.37 (2H, d, J = 8.5 Hz, PMB CH_{Ar} x 2), 6.98 (2H, d, J = 8.5 Hz, PMB CH_{Ar} x 2), 6.61 (2H, m, Dmb CH_{Ar}), 4.40 (1H, m, Dmb CH), 3.94 (9H, m, OMe x 3), 3.78 (2H, 2 x s, PMB CH_2), 3.02 (0.4H, d, J = 4.4 Hz, S-CH_2), 2.96 (0.6H, J = 4.4 Hz, S-CH_2), 2.78 (0.6H, J = 8.6 Hz, S-CH_2), 2.74 (0.4H, J = 8.6 Hz, S-CH_2), 2.00 (2H, bs, NH_2). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 159.8, 158.4, 157.5, 130.4 (qC x 4), 129.8 (CH_{Ar} x 2), 127.4 (CH_{Ar}), 124.5 (qC), 113.6 (CH_{Ar} x 2),, 103.8, 98.5 (CH_{Ar} x 2), 55.2, 55.1, 55.0 (CH_3 x 3), 49.4 (CH), 39.0, 35.5 (CH_2 x 2). HRMS (FAB): found 333.1367. $C_{18}H_{23}NO_3S$ $[M]^+$ requires 333.1399.

Typical procedure for preparation of acid fluorides



34a: R = H (99 %)
34b: R = Me (99 %)
34c: R = $(CH_2)_4NHBoc$ (100 %)

The Fmoc-protected amino acid (500 mg) was dissolved in dry DCM (25.0 ml) under argon and pyridine (1.1 eq.) was added. The reaction mixture was cooled to -10 °C and cyanuric fluoride (5.0 eq.) was added dropwise. The reaction mixture was allowed to warm to room temperature and stirring was continued for 3 hours. A further portion of DCM (10.0 ml) was added, followed by H_2O (5.0 ml). The reaction mixture was then filtered and the organic phase was collected and washed with H_2O (2 x 15.0 ml), then dried over $MgSO_4$ and evaporated to dryness in vacuo to afford the acid fluorides as white solids which were used without further purification.

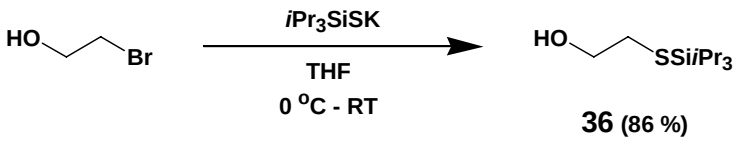
***N*-9-fluorenylmethoxycarbonyl-glycine acid fluoride 34a**: 99 %. m.p. 132-133 °C (literature 134-135 °C, reference 199). IR (nujol) ν_{max} 3334, 1844, 1679 cm^{-1} (literature 1850 cm^{-1}, reference 199). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.69 (2H, J = 7.2 Hz, Fmoc CH_{Ar} x 2), 7.49 (2H, d, J = 7.3 Hz, Fmoc CH_{Ar} x 2), 7.33-7.20 (4H, m, Fmoc CH_{Ar} x 4), 5.19 (1H, bs, NH), 4.38 (2H, d, J = 6.9 Hz, Fmoc CH_2), 4.60-4.04 (3H, m, Cα-H_2 + Fmoc CH). ^{19}F-NMR (235.3 MHz, $CDCl_3$): δ = 30.7. LRMS (ESI): found 299.9. $C_{17}H_{15}FNO_3$ $[MH]^+$ requires 300.10.

***N*-9-fluorenylmethoxycarbonyl-alanine acid fluoride 34b**: 99 %. m.p. 106-107 °C (literature 108 °C, reference 199). IR (nujol) ν_{max} 3319, 1846, 1686 cm^{-1} (literature 1850

cm^{-1}, reference 199). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.69 (2H, J = 7.2 Hz, Fmoc CH_{Ar} x 2), 7.50 (2H, d, J = 6.9 Hz, Fmoc CH_{Ar} x 2), 7.38-7.20 (4H, m, Fmoc CH_{Ar} x 4), 5.12 (1H, d, J = 7.6 Hz, NH), 4.51-4.30 (3H, m, Fmoc CH_2 + Cα-H), 4.15 (1H, t, J = 6.6 Hz, Fmoc CH) 1.46 (3H, d, J = 7.4 Hz, Cα-CH_3). ^{19}F-NMR (235.3 MHz, $CDCl_3$): δ = 26.7. LRMS (ESI): found 314.1. $C_{18}H_{17}FNO_3$ $[MH]^+$ requires 314.12.

***N*-9-fluorenylmethoxycarbonyl-lysine(*N*-*t*butoxycarbonyl) acid fluoride 34c**: 100 %. m.p. 121-122 °C (literature 120-121 °C, reference 199). IR (nujol) ν_{max} 3354, 1854, 1682 cm^{-1} (literature 1850 cm^{-1}, reference 199). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 7.69 (2H, J = 7.3 Hz, Fmoc CH_{Ar} x 2), 7.52 (2H, d, J = 7.2 Hz, Fmoc CH_{Ar} x 2), 7.37-7.20 (4H, m, Fmoc CH_{Ar} x 4), 5.53 (1H, d, J = 7.6 Hz, NH), 4.54 (1H, bm, NH), 4.46-4.28 (3H, m, Fmoc CH_2 + Cα-H), 4.14 (1H, t, J = 6.8 Hz, Fmoc CH), 3.06 (2H, m, CH_2), 1.90-1.72 (2H, m, CH_2) 1.44-1.39 (2H, m, CH_2), 1.36 (9H, s, Boc *t*Bu). ^{19}F-NMR (235.3 MHz, $CDCl_3$): δ = 28.5. LRMS (ESI): found 471.1. $C_{26}H_{32}FN_2O_5$ $[MH]^+$ requires 471.23.

2-hydroxy-(S-triisopropylsilyl)-ethanethiol 36[173]



36 (86 %)

Potassium hydride (35 % suspension in mineral oil, 268 mg, 2.40 mmol) was washed with anhydrous pentane (3 x 5.0 ml), then suspended in anhydrous pentane (5.0 ml) and the reaction mixture was cooled to 0 °C. Triisopropylsilyl thiol (0.50 ml, 2.30 mmol) was added dropwise to the suspension and stirring was continued at 0 °C for 2 hours. The reaction mixture was evaporated to dryness in vacuo and recrystallised from toluene to

afford the potassium triisoproylsilylthiolate salt as a white solid (520 mg, 2.28 mmol, 99 %) which was then suspended in anhydrous THF under argon. The suspension was cooled to 0 ^{o}C and bromoethanol (200 µl, 2.80 mmol) was added dropwise. Stirring was continued for 15 minutes at 0 ^{o}C, then for 5 hours at room temperature. The resulting white precipitate was removed by filtration and washed with anhydrous THF (15.0 ml). The combined filtrate was concentrated in vacuo and purified by column chromatography over silica (EtOAc/petroleum ether bp 40-60 ^{o}C, 15:85) to afford **36** as a clear oil (464 mg, 1.98 mmol, 86 % based on starting triisopropylsilyl thiol). ^{1}H-NMR (250 MHz, $CDCl_3$): δ = 3.66 (2H, t, J = 5.9 Hz, CH_2), 2.69 (2H, t, J = 6.1 Hz, CH_2), 1.74 (1H, bs, OH), 1.26-1.11 (3H, m, *i*Pr CH x 3), 1.03 (18H, s, *i*Pr CH_3 x 6). ^{13}C-NMR (62.8 MHz, $CDCl_3$): δ = 63.8, 29.6 (CH_2 x 2), 18.9 (CH_3 x 6), 13.1 (CH x 3). LRMS (ESI): found 216.9. $C_{11}H_{25}SSi$ $[M\text{-}OH]^+$ requires 217.14. ^{1}H and ^{13}C spectroscopic data in good agreement with literature (reference 173).

Peptide Synthesis

Except where noted, manual peptide synthesis was carried out on a 0.05 mmol scale using 10 equivalents of Fmoc-amino acid per coupling and HOBt/HBTU coupling reagents. The Fmoc amino acid was dissolved in DMF (1.0 ml) and DIPEA (150 µl, 0.85 mmol) and 1.0 ml of an HOBt/HBTU solution in DMF (0.5 M concentration for each reagent) were added with stirring. The resulting solution was added to the peptide synthesis tube containing the resin and agitated for an average of 4 hours. Fmoc removal was carried out by treatment of the resin with a 20 % v/v solution of piperidine in DMF for 20

minutes. Reaction progress was monitored by the Kaiser ninhydrin test and LC-MS analysis. 5 equivalents of auxiliary-amino acid cassettes and coupling reagents were used for the final coupling. Resins were washed exhaustively with DMF and then with DCM after each coupling or deprotection step. All resins, amino acids and coupling reagents were purchased from Novabiochem.

Peptides were cleaved from the solid phase with concomitant side-chain deprotection by treatment with 95 % TFA v/v, 2.5 % EDT v/v, 2.5 % H_2O v/v for 3 hours. The resin was removed by filtration and the crude peptide was precipitated from ice-cold ether and collected by centrifugation at 5000 rpm for 15 mins. The crude peptide was then dissolved in the minimum volume of MeCN/H_2O (0:100-20:80) and purified by semi-preparative HPLC. Fractions containing the product were lyophilized to afford the peptides as fluffy white solids.

UV Fmoc analysis was conducted as described in the Novabiochem synthesis notes. Approximately 1 μmol of resin with respect to Fmoc was weighed into each of two 10 mm silica UV cells. Freshly prepared 20 % v/v piperidine in DMF (3.0 ml) was added to each cell and also to a third empty cell. The resin mixture in each cell was gently agitated with a pasteur pipette for 3 minutes. The UV absorbance of each of the two cells containing the resin mixture was measured at 290 nm, using the cell containing only the piperidine solution as a reference. The formula loading = (Abs_{sample} - Abs_{ref})/(1.65 x mg resin) was then used to obtain an estimate of the resin loading (mmol/g).

Auxiliary-peptides were synthesized on NovaSyn®TGT resin preloaded with serine (loading = 0.22 mmol/g). Auxiliary-peptides **24a** and **31a** have the sequence Aux-GSETVEEELGKIIEGFVTGAEDIISGASRITKS, corresponding to residues 100-132 of GlyCAM-1. Auxiliary-peptides **24b**, **24c**, **31b** and **31c** are shorter versions of the same sequence as detailed in the text. Auxiliary-glycopeptide **32** has the sequence Aux-GSSQLEET(GalNAc$(OAc)_3$)S, corresponding to GlyCAM-1 residues 76-83.

Peptide benzyl thioesters **33a – 33d** were synthesized on 4-sulfamylbutyryl-AM resin (loading 1.1 mmol/g). Peptide alkyl thioesters **37a – 37c** were synthesized on Rink amide AM resin (loading 0.62 mmol/g). Benzyl thioester **33a** was also synthesized on this resin as detailed in **Scheme 35**.

***N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzyl glycine-(GlyCAM-1 101-132) 23a**: 7 % isolated yield. m/z (ESI) calculated: 3756.2, observed: 1878.5 $[M]^{2+}$, 1712.4 $[M\text{-}Aux]^{2+}$, 1252.4 $[M]^{3+}$, 1141.6 $[M\text{-}Aux]^{3+}$.

***N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzyl glycine-(GlyCAM-1 110-132) 23b**: 10 % isolated yield. m/z (ESI) calculated: 2782.2, observed: 1392.0 $[MH_2]^{2+}$, 1225.8 $[M\text{-}AuxH_2]^{2+}$.

***N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzyl alanine-(GlyCAM-1 116-132) 23c**: 10 % isolated yield. m/z (ESI) calculated: 2108.4, observed: 2108.2 $[M]^{+}$, 1776.6 $[M\text{-}Aux]^{+}$, 1055.0 $[MH_2]^{2+}$, 888.9 $[M\text{-}AuxH_2]^{2+}$.

***N*-4,5,6-trimethoxy-2-(2-nitrobenzylthio)benzyl glycine-(GlyCAM-1 110-132) 23d**: m/z (ESI) calculated: 2797.1, observed: 1399.1 $[MH_2]^{2+}$, 1225.8 $[M\text{-}AuxH_2]^{2+}$.

***Nα*-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)-glycine-(GlyCAM-1 101-132) 31a**: 10 % isolated yield. m/z (ESI) calculated: 3620.0, observed: 1810.0 $[M]^{2+}$, 1712.2 $[M\text{-}Aux]^{2+}$, 1207.6 $[M]^{3+}$, 1141.8 $[M\text{-}AuxH_3]^{3+}$.

***Nα*-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)-glycine-(GlyCAM-1 110-132) 31b**: 11 % isolated yield. m/z (ESI) calculated: 2646.0, observed: 1323.9 $[MH_2]^{2+}$, 1225.9 $[M\text{-}AuxH_2]^{2+}$, 883.2 $[MH_3]^{3+}$, 817.7 $[M\text{-}AuxH_3]^{3+}$.

***Nα*-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)-alanine-(GlyCAM-1 116-132) 31c**: 13 % isolated yield. m/z (ESI) calculated: 1972.2, observed: 1972.5 $[M]^{+}$, 1776.4 $[M\text{-}Aux]^{+}$, 987.0 $[MH_2]^{2+}$, 889.0 $[M\text{-}AuxH_2]^{2+}$.

Typical procedure for PMB deprotection of *N*-4,5,6-trimethoxy-2-(4-methoxybenzylthio)benzyl peptides 23a-23c:

The *S*-PMB protected peptide (1.0 mg) was placed in an eppendorf tube and dissolved in 10 % v/v acetic acid (0.50 ml) at 0 ^{o}C. Mercury(II)acetate (1.0 mg) was added and the reaction mixture was shaken (250 rpm) at room temperature under argon for the times shown in the text. DTT was added to a final concentration of 5 % w/v and the reaction mixture was agitated for a further 1 hour. The thick white precipitate was removed by centrifugation at 13000 rpm for 10 mins. The remaining solution was purified directly by semi-preparative HPLC and fractions lyophilized to afford the *N*-4,5,6-trimethoxy-2-mercapto-benzyl peptides **24a - 24c** as fluffy white solids.

***N*-4,5,6-trimethoxy-2-mercapto-benzyl glycine-(GlyCAM-1 101-132) 24a**: 69 % isolated yield. m/z (ESI) calculated: 3636.0, observed: 1818.3 $[M]^{2+}$, 1712.5$[M\text{-}AuxH_2]^{2+}$, 1212.3 $[M]^{3+}$, 3425.0 $[M\text{-}AuxH_3]^{3+}$.

***N*-4,5,6-trimethoxy-2-mercapto-benzyl glycine-(GlyCAM-1 110-132) 24b**: 87 % isolated yield. m/z (ESI) calculated: 2662.02, observed: 1331.8 $[MH_2]^{2+}$, 1225.8 $[M\text{-}AuxH_2]^{2+}$, 888.4 $[MH_3]^{3+}$, 817.6 $[M\text{-}AuxH_3]^{3+}$.

***N*-4,5,6-trimethoxy-2-mercapto-benzyl alanine-(GlyCAM-1 116-132) 24c**: 71 % isolated yield. m/z (ESI) calculated: 1988.2, observed: 1988.7 $[M]^{+}$, 995.1 $[MH_2]^{2+}$, 889.0 $[M\text{-}AuxH_2]^{2+}$.

***Nα*-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl) glycine (GlyCAM-1 74-83) glycopeptide 32**:

S-PMB protected glycopeptide **31d** (1.0 mg) was dissolved in a minimum volume (100 μl) of ice-cold TFA and treated with mercury(II)acetate (1.0 mg). The reaction mixture was allowed to stand under argon at 0 ^{o}C for 10 mins then diluted to 10 % v/v TFA (1.0 ml) with H_2O. DTT was then added to a final concentration of 5 % w/v. The mixture was shaken (250 rpm) at room temperature for 30 minutes and the thick white precipitate was removed by centrifugation at 13000 rpm for 10 minutes. The remaining solution was purified directly by semi-preparative HPLC and fractions lyophilized to afford the thiol as a fluffy white solid (quantitative yield). m/z (ESI) calculated: 1462.5, observed: 1462.7 $[M]^{+}$, 1266.9 $[M\text{-}Aux]^{+}$, 937.3 $[M\text{-}Aux\text{-}(AcO)_3GalNAc]^{+}$.

Typical procedure for synthesis of benzyl thioesters 33a-33d.

Initial coupling with DIPCDI: 45 mg (0.05 mmol) of 4-sulfamylbutyryl-AM resin (loading 1.1 mmol/g) was preswelled in DCM for 1 hour. The Fmoc-protected *C*-terminal amino acid (0.20 mmol) and *N*-methyl imidazole (16 µl, 0.20 mmol) were dissolved in a 4:1 v/v solution of DCM/DMF and DIPCDI (31 µl, 0.20 mmol) was added. The reagents were mixed and then added to the resin. The tube containing the reaction mixture was agitated for 18 hours, after which time the tube was drained and the resin was washed exhaustively with DMF and DCM. This coupling step was then repeated as above. Subsequent peptide elongation was carried out exactly as for auxiliary-peptides. The *N*-terminal residue of each peptide was coupled as the Boc-protected amino acid to avoid a further Fmoc deprotection step. The resin was transferred to a glass vial and a solution of iodoacetonitrile (90 µl, 1.25 mmol, prefiltered through a plug of basic alumina) and DIPEA (95 µl, 0.55 mmol) in NMP (1.0 ml) was added and the reaction mixture was agitated in the absence of light for 12-18 hours. The resin was removed by filtration and washed exhaustively with NMP, DCM and THF, then added to a glass vial. THF (1.5 ml) was added, followed by benzyl mercaptan (205 µl, 1.75 mmol) and NaSPh (3.0 mg, 0.02 mmol) and the reaction mixture was stirred for 18-24 hours. The resin was removed by filtration and washed with DCM (3 x 5.0 ml). The combined filtrate was concentrated in vacuo and then treated with 95 % v/v TFA, 2.5 % v/v EDT, 2.5 % v/v H_2O for 3 hours. The crude product was precipitated with ice cold ether and collected by centrifugation at 5000 rpm for 15 minutes, then purified by semi-preparative HPLC as for auxiliary-peptides.

***C*-terminal alanine benzyl thioester 33a**: < 1 % isolated yield. m/z (ESI) calculated: 653.7, observed: 654.6 $[MH]^+$, 676.2 $[MNa]^+$, 1307.4 $[2M]^+$.

***C*-terminal phenylalanine benzyl thioester 33b**: 5 % isolated yield. m/z (ESI) calculated: 729.8, observed: 730.4 $[M]^+$.

***C*-terminal lysine benzyl thioester 33c**: 7 % isolated yield. m/z (ESI) calculated: 608.8, observed: 609.1 $[M]^+$.

Initial coupling with acid fluoride: 90 mg (0.10 mmol) of 4-sulfamylbutyryl-AM resin (loading 1.1 mmol/g) was suspended in anhydrous DCM (300 µl) in a glass vial. DIPEA (35 µl, 0.20 mmol) was added, followed by the Fmoc-protected acid fluoride **34a** or **34b** (0.30 mmol), dissolved in anhydrous DCM (1.20 ml). The vial contents were stirred for 2 hours, then the resin was removed by filtration and washed exhaustively with DCM. A sample of the resin was subjected to UV Fmoc analysis and a loading of 0.84 mmol/g (76 %) from the reaction with glycine acid fluoride **34a** and 0.81 mmol/g (74 %) from the reaction with alanine acid fluoride **34b** was observed. Peptide elongation and the activation and cleavage steps were conducted exactly as above to afford alanine thioester **33a** in 2 % yield and glycine thioester **33d**.

***C*-terminal glycine benzyl thioester 33d**: 7 % isolated yield. m/z (ESI) calculated: 507.7, observed: 508.4 $[MH]^+$.

Double linker resin 35

161 mg (0.10 mmol) of Fmoc-Rink amide AM resin (loading 0.62 mmol/g) was preswelled in DCM for 1 hour, then washed with DMF and subjected to standard Fmoc deprotection. Fmoc-protected phenylalanine (194 mg, 0.50 mmol) was coupled using standard HOBt/HBTU coupling in the presence of DIPEA with 5 equivalents of coupling reagents and Fmoc deprotection was again carried out. 3-carboxypropanesulfonamide (50 mg, 0.30 mmol) and HOBt (41 mg, 0.30 mmol) were dissolved in anhydrous DMF (2.0 ml) and DIPCDI (47 µl, 0.30 mmol) was added. These reagents were mixed and then added to the resin and agitated for 5 hours. The *C*-terminal residue of the target thioester was coupled as the acid fluoride (0.30 mmol) in the presence of DIPEA (52 µl, 0.30 mmol) in DCM (1.5 ml) for 4 hours. Samples of the resin were then subjected to UV Fmoc analysis and loadings of 0.47 mmol/g (76 %) for glycine, 0.33 mmol/g (54 %) for alanine and 0.36 mmol/g (58 %) for lysine were observed. Peptide elongation was carried out exactly as before. The resin bearing the *C*-terminal alanine peptide was split in two and half was subjected to the previously used iodoacetonitrile activation and benzyl mercaptan cleavage procedures as discussed in the text. The remainder of this resin was preswelled in the minimum volume of anhydrous THF for 1 hour, then drained and washed with THF and placed in a glass vial. Alcohol **36** (82 mg, 0.35 mmol) and PPh_3 (92 mg, 0.35 mmol) were dissolved in anhydrous THF and added to the resin and the reaction mixture was cooled to 0 ^{o}C. DEAD (71 µl, 0.45 mmol) was added dropwise and the reaction mixture was stirred overnight, over which time the temperature was allowed to rise to room temperature. The resin was removed by filtration and washed repeatedly with THF, DCM and then with further portions of THF. If necessary, as

determined by LC-MS analysis of a the crude residue obtained by subjection of a sample of the resin to TFA cleavage, the Mitsunobu alkylation was repeated as above, otherwise the resin was subjected to TIPS deprotection. A solution of TBAF and AcOH in THF was made up by mixing 1M TBAF in THF (0.50 ml) with 0.2 M AcOH in THF (5.0 ml). 3.0 ml of the resulting solution was added to the resin and stirred for 30 minutes, after which time the resin was removed by filtration and washed (THF, DCM, Et_2O). The resin was then treated with 95 % v/v TFA, 2.5 % v/v EDT, 2.5 % v/v H_2O v/v for 3 hours. The resin was removed by filtration and the crude peptide was collected by precipitation of the filtrate in ice-cold Et_2O/hexane 1:1 v/v and centrifugation and purified by semi-preparative HPLC as before to afford alkyl thioester peptide **37a**. *C*-terminal glycine and lysine alkyl thioesters **37b** and **37c** were produced by the same procedure, scaling up as appropriate.

***C*-terminal alanine alkyl thioester 37a**: 11 % isolated yield. m/z (ESI) calculated: 903.0, observed: 903.7 $(MH]^+$.

***C*-terminal glycine alkyl thioester 37b**: 12 % isolated yield. m/z (ESI) calculated: 756.9, observed: 757.7 $[MH]^+$.

***C*-terminal lysine alkyl thioester 37c**: 18 % isolated yield. m/z (ESI) calculated: 858.1, observed: 858.6 $[M]^+$.

Typical procedure for thiol-free ligation reactions:

The auxiliary-peptide was dissolved in ligation buffer (6 M guanidine hydrochloride, 200 mM sodium phosphate, pH = 8.0) to a concentration of 8 mM. 200 μl of this solution

was added under a stream of argon to an eppendorf tube containing 200 µl of a similar 8 mM solution of the thioester peptide in ligation buffer, followed by 8 µl of a 1 M TCEP solution (final TCEP concentration 20 mM). The reaction mixture was shaken under argon at room temperature for 48 – 72 hours and monitored by LC-MS. 8.0 mg MESNa was added to give a final concentration of 2 % w/v and agitation was continued for a further 6 hours, after which the reaction mixture was purified directly by semi-preparative reverse phase HPLC (in the initial ligation studies in the presence of thiol, the same amount of MESNa was added after the TCEP solution at the start of the ligation).

Gly-Gly (4,5,6-trimethoxy-2-mercaptobenzyl) ligation product 39a: 54 % isolated yield. m/z (ESI) calculated: 3045.5, observed: 1523.6 $[MH_2]^{2+}$, 1016.1 $[MH_3]^{3+}$.

Gly-Gly [*N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)] ligation product 39b: 67 % isolated yield. m/z (ESI) calculated: 3029.5, observed: 1515.5 $[MH_2]^{2+}$, 1010.9 $[MH_3]^{3+}$.

Gly-Gly [*N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)] glycopeptide ligation product 42: 48 % isolated yield. m/z (ESI) calculated: 1845.9, observed: 1845.9 $[M]^{+}$, 1650.1 $[M\text{-aux}]^{+}$, 1517.3 $[M\text{-aux-}(OAc)_3GalNAc]^{+}$, 923.8 $[MH_2]^{2+}$.

Ala-Gly [*N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)] ligation product 46: 70 % isolated yield. m/z (ESI) calculated: 3175.5, observed: 1588.7 $[MH_2]^{2+}$, 1059.6 $[MH_3]^{3+}$.

Lys-Gly [*N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)] ligation product 48: 24 % isolated yield. m/z (ESI) calculated: 4103.6, observed: 1369.1 $[MH_3]^{3+}$, 1026.8 $[MH_4]^{4+}$.

Gly-Ala (4,5,6-trimethoxy-2-mercaptobenzyl) ligation product 50a: 42 % isolated yield. m/z (ESI) calculated: 2371.7, observed: 1186.7 $[MH_2]^{2+}$, 791.6 $[MH_3]^{3+}$.

Gly-Ala [*N*α-(1-(2,4-dimethoxyphenyl)-2-mercaptoethyl)] ligation product 50b: 48 % isolated yield. m/z (ESI) calculated: 2355.7, observed: 1178.8 $[MH_2]^{2+}$, 786.6 $[MH_3]^{3+}$.

Typical procedure for auxiliary removal:

The purified ligation products were placed in an eppendorf tube under argon and dissolved in 95 % v/v TFA, 2.5 % v/v EDT, 2.5 % v/v H_2O v/v and allowed to stand with occasional agitation at room temperature for 3 hours. After this time the bulk of the TFA was evaporated under a stream of argon and the crude product was collected by precipitation in ice-cold Et_2O and centrifugation and purified by semi-preparative HPLC as before.

Gly-Gly ligation product 40: 58 % isolated yield from **39a**, 100 % isolated yield from **39b**. m/z (ESI) calculated: 2833.2, observed: 1417.4 $[MH_2]^{2+}$, 945.6 $[MH_3]^{3+}$.

Gly-Gly glycopeptide ligation product 43: m/z (ESI) calculated: 1649.7, observed: 1650.1 $[M]^{+}$, 1320.9 $[M\text{-}(OAc)_3GalNAc]^{+}$, 825.6 $[MH_2]^{2+}$.

Ala-Gly ligation product 47: 100 % isolated yield from **46**. m/z (ESI) calculated: 2979.3, observed: 1490.5 $[MH_2]^{2+}$, 994.3 $[MH_3]^{3+}$.

Lys-Gly ligation product 49: 100 % isolated yield from **48**. m/z (ESI) calculated: 3907.3, observed: 1954.6 $[MH_2]^{2+}$, 1303.6 $[MH_3]^{3+}$, 977.9 $[MH_4]^{4+}$.

Gly-Ala ligation product 51: 60 % isolated yield from **50a**, 100 % isolated yield from **50b**. m/z (ESI) calculated: 2159.4, observed: 1080.7 $[MH_2]^{2+}$, 720.9 $[MH_3]^{3+}$.

Deprotected Gly-Gly glycopeptide ligation product 44

The purified ligation product **43** was placed in an eppendorf tube under argon and dissolved in sodium phosphate buffer pH 8.0 containing 2 % hydrazine hydrate v/v and

allowed to stand with occasional agitation at room temperature for 4 hours. Complete reaction was confirmed by LC-MS analysis. m/z (ESI) calculated: 1523.6, observed: 1524.0 $[M]^+$, 1320.8 $[M\text{-}(OAc)_3GalNAc]^+$, 762.6 $[MH_2]^{2+}$.

Acknowledgements

This project was a collaborative work and the following practical contributions by Dr Derek Macmillan are gratefully acknowledged: preparation of ONB-protected auxiliary-peptide **23d**; preparation of PMB-protected auxiliary-glycopeptide **31d** and development of deprotection conditions to give **32**; preparation of thioesters **33e** and **33f**; preparation of glycopeptide ligation product **42** and deprotection to give **44**; and initial ligation studies at Leu-Gly and Ser-Gly junctions in the presence of MESNa.

23d was prepared as described for auxiliary-peptides, with use of cassette **11b** in the final coupling step. Glycopeptide **31d** was prepared as described for auxiliary-peptides, with use of the Fmoc-Ser[$(OAc)_3$GalNAc]-OH cassette, prepared and incorporated as described in reference 196, and with the use of cassette **13a** in the final coupling step. Protecting group and auxiliary removal steps were performed as described in the experimental section.

In addition to his practical contribution, I would like to thank Dr Derek Macmillan for all his support, encouragement, advice and patience as my supervisor. Thanks also to all past and current members of the Macmillan research group and all those in lab 34, and the technical support staff in the School of Chemistry, especially Robert Smith in mass spectrometry and John Millar in NMR, for their assistance.

Financial support was provided by the BBSRC and Cyclacel.

References

1. Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1999.
2. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683-720.
3. Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364-2369.
4. Spiro, R. G. *Glycobiology* **2002**, *12*, 43-56.
5. Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357.
6. Davis, B. J. *Chem. Rev.* **2002**, *102*, 579-601.
7. O'Connor, S. E.; Imperiali, B. *Chemistry & Biology* **1996**, *3*, 803-812.
8. Pratt, M. R.; Bertozzi, C. R. *Chem. Soc. Rev.* **2005**, *34*, 58-68.
9. Rush, R. S.; Derby, P. L.; Smith, D. M.; Merry, C.; Rogers, G.; Rohde, M. F.; Katta, V. *Anal. Chem.* **1995**, *67*, 1442-1452.
10. Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17-22.
11. Taylor, C. M. *Tetrahedron* **1998**, *54*, 11317-11362.
12. Shriver, Z.; Raguram, S.; Saisekharan, R. *Nat. Rev. Drug Disc.* **2004**, *3*, 863-873.
13. Weerapana, E.; Imperiali, B. *Glycobiology* **2006**, *16*, 91-101.
14. Shao, N.; Guo, Z.-W. *J. Org. Chem.* **2003**, *68*, 9003-9011.
15. Shao, N.; Xue, J.; Guo, Z.-W. *Angew. Chem. Int. Edn.* **2004**, *43*, 1569-1573.
16. Goldwasser, E. *Blood Cells* **1984**, *10*, 147-164.
17. Fukada, M. N.; Sasaki, H.; Lopez, L.; Fukada, M. *Blood* **1989**, *73*, 84-89.
18. Smith, A. E.; Helenius, A. *Science* **2004**, *304*, 237-242.
19. Ellgaard, L.; Helenius, A. *Curr. Opin. Cell Biol.* **2001**, *13*, 431-437.
20. Taylor, S. C.; Ferguson, A. D.; Bergeron, J. J. M.; Thomas, D. Y. *Nat. Struct. Mol. Biol.* **2004**, *11*, 128-134.
21. Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 151-208.
22. Pratt, M. R.; Hang, H. C.; Ten Hagen, K. G.; Rarick, J.; Gerken, T. A.; Tabak, L. A.; Bertozzi, C. R. *Chem. Biol.*. **2004**, *11*, 1009-1016.
23. Hang, H. C.; Bertozzi, C. R. *Bioorg. Med. Chem.* **2005**, *13*, 5021-5034.
24. Fritz, T. A.; Hurley, J. A.; Trinh, L.-B.; Shiloach, J.; Tabak, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15307-15312.
25. Brocke, C.; Kunz, H. *Bioorg. Med. Chem.* **2002**, *10*, 3085-3112.
26. Leppanen, A.; White, S. P.; Helin, J.; McEver, R. P.; Cummings, R. D. *J. Biol. Chem.* **2000**, *275*, 39569-39578.
27. Tseng, P.-H.; Jiaang, W.-T.; Chang, M.-Y.; Chen, S.-T. *Chem. Eur. J.* **2001**, *7*, 585-590.
28. Danishefsky, S. J.; Allen, J. R. *Angew. Chem. Int. Edn.* **2000**, *39*, 836-863.
29. Witte, K.; Sears, P.; Wong, C.-H. *J. Am. Chem. Soc.* **1997**, *119*, 2114-2118.
30. Takegawa, K.; Tabuchi, M.; Yamaguchi, S.; Kondo, A.; Kato, I.; Iwahara, S. *J. Biol. Chem.* **1995**, *270*, 3094-3099.
31. Warner, T. G. *Glycobiology* **1999**, *9*, 841-850.
32. Gerngross, T. U. *Nat. Biotechnol.* **2004**, *22*, 1409-1414.
33. Choi, B.-K.; Bobrowicz, P.; Davidson, R. C.; Hamilton, S. R.; Kung, D. H.; Li, H.; Miele, R. G.; Nett, J. H.; Wildt, S.; Gerngross, T. U. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5022-5027.

34. Zhang, Z.; Gildersleeve, J.; Yang, Y.-Y.; Xu, R.; Loo, J. A.; Uryu, S.; Wong, C.-H.; Schultz, P. G. *Science* **2004**, *303*, 371-373.
35. Xu, R.; Hanson, S. R.; Zhang, Z.; Yang, Y.-Y.; Schultz, P. G.; Wong, C.-H. *J. Am. Chem. Soc.* **2004**, *126*, 15654-15655.
36. Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.
37. Cammish, L. E.; Kates, S. A. In *Fmoc Solid-phase Peptide Synthesis: A Practical Approach*; Chan, W. C., White, P. D., Eds.; Oxford University Press: Oxford, 2000, p 227.
38. Kimmerlin, T.; Seebach, D. *J. Pept. Res.* **2005**, *65*, 229-260.
39. Montalbetti, C. A. G. N.; Falque, V. *Tetrahedron* **2005**, *61*, 10827-10852.
40. Ito, Y.; Ohnishi, Y.; Ogawa, T.; Nakahara, Y. *Synlett.* **1998**, 1102-1104.
41. Kim, J.-H.; Yang, H.; Boons, G.-J. *Angew. Chem. Int. Edn.* **2005**, *44*, 947-949.
42. Nicolaou, K. C.; Mitchell, H. J. *Angew. Chem. Int. Edn.* **2001**, *40*, 1576-1624.
43. Sears, P.; Wong, C.-H. *Science* **2001**, *291*, 2344-2350.
44. Davis, B. J. *J. Chem. Soc. Perkin Trans. 1* **2000**, 2137-2160.
45. Koenigs, W.; Knorr, E. *Ber. Dtsch. Chem. Ges.* **1901**, *34*, 957.
46. Holemann, A.; Seeberger, P. H. *Curr. Opin. Biotechnol.* **2004**, *15*, 615-622.
47. Macmillan, D.; Daines, A. M. *Curr. Med. Chem.* **2003**, *10*, 2733-2773.
48. Seeberger, P. H. *Chem. Commun.* **2003**, 1115-1121.
49. Ratner, D. M.; Swanson, E. R.; Seeberger, P. H. *Org. Lett.* **2003**, *5*, 4717-4720.
50. Zhang, Z.; Ollmann, I. R.; Ye, X. S.; Wischnat, R.; Baasov, T.; Wong, C.-H. *J. Am. Chem. Soc.* **1999**, *121*, 734-753.
51. Burkhart, F.; Zhang, Z.; Wacowich-Sgarbi, S.; Wong, C.-H. *Angew. Chem. Int. Edn.* **2001**, *41*, 4087-4090.
52. Wong, C.-H. *J. Org. Chem* **2005**, *70*, 4219-4225.
53. Liu, L.; Bennett, C. S.; Wong, C.-H. *Chem. Commun.* **2006**.
54. Li, B.; Zeng, Y.; Hauser, S.; Song, H.; Wang, L.-X. *J. Am. Chem. Soc.* **2005**, *127*, 9692-9693.
55. Li, H.; Li, B.; Song, H.; Breydo, L.; Baskakov, I.; Wang, L.-X. *J. Org. Chem.* **2005**, *70*, 9990-9996.
56. Yamamoto, N.; Sakakibara, T.; Kajihara, Y. *Tetrahedron Lett.* **2004**, *45*, 3287-3290.
57. Winterfield, G. A.; Schmidt, R. R. *Angew. Chem. Int. Edn.* **2001**, *40*, 2654-2657.
58. Kunz, H.; Unverzagt, C. *Angew. Chem. Int. Edn.* **1988**, *27*, 1697-1699.
59. Sjolin, P.; Elofsson, M.; Kihlberg, J. *J. Org. Chem.* **1996**, *61*, 560-565.
60. Koeller, K. M.; Smith, M. E. B.; Huang, R.-F.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *112*, 4241-4242.
61. Matsushita, T.; Hinou, H.; Fumoto, M.; Kurogochi, M.; Fujitani, N.; Shimizu, H.; Nishimura, S.-I. *J. Org. Chem.* **2006**, *71*, 3051-3063
62. Mezzato, S.; Schaffrath, M.; Unverzagt, C. *Angew. Chem. Int. Edn.* **2005**, *44*, 1650-1654.
63. Kajihara, Y.; Suzuki, Y.; Yamamoto, N.; Sasaki, K.; Sakakibara, T.; Juneja, L. R. *Chem. Eur. J.* **2004**, *10*, 971-985.
64. Hojo, H.; Haginoya, E.; Matsumoto, Y.; Nakahara, Y.; Nabeshima, K.; Toole, B. P.; Watanabe, Y. *Tetrahedron Lett.* **2003**, *44*, 2961-2964.
65. Yamamoto, N.; Takayanagi, A.; Sakakibara, T.; Dawson, P. E.; Kajihara, Y. *Tetrahedron Lett.* **2006**, *47*, 1341-1346.
66. Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 6576-6578.

67. Dudkin, V. Y.; Miller, J. S.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 736-738.
68. Mandal, M.; Dudkin, V. Y.; Geng, X.; Danishefsky, S. J. *Angew. Chem. Int. Edn.* **2004**, *43*, 2557-2561.
69. Cohen-Anisfeld, S. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1993**, *115*, 10531-10537.
70. Wang, Z.-G.; Warren, J. D.; Dudkin, V. Y.; Zhang, X.; Iserloh, U.; Visser, M.; Eckhardt, M.; Seeberger, P. H.; Danishefsky, S. J. *Tetrahedron* **2006**, *62*, 4954-4978.
71. Hernzer, H.; Reipen, T.; Schultz, M.; Kunz, H. *Chem. Rev.* **2000**, *100*, 4495-4537.
72. Kaneshiro, C. M.; Michael, K. *Angew. Chem. Int. Edn.* **2006**, *45*, 1077-1081.
73. Danishefsky, S. J.; Hu, S.; Cirillo, P. E.; Eckhardt, M.; Seeberger, P. H. *Chem. Eur. J.* **1997**, *3*, 1617-1628.
74. Miller, J. S.; Dudkin, V. Y.; Lyon, G. J.; Muir, T. W.; Danishefsky, S. J. *Angew. Chem. Int. Edn.* **2003**, *42*, 431-434.
75. Geng, X.; Dudkin, V. Y.; Mandal, M.; Danishefsky, S. J. *Angew. Chem. Int. Edn.* **2004**, *43*, 2562-2565.
76. Gunther, W.; Kunz, H. *Angew. Chem. Int. Edn.* **1990**, *29*, 1050-1051.
77. Handlon, A. L.; Fraser-Reid, B. *J. Am. Chem. Soc.* **1993**, *115*, 3796-3797.
78. Kahne, D.; Walker, S.; Cheng, Y.; Vanengen, D. *J. Am. Chem. Soc.* **1989**, *111*, 6881-6882.
79. He, Y.; Hinklin, R. J.; Chang, J.; Kiessling, L. L. *Org. Lett.* **2004**, *6*, 4479-4482.
80. Doores, K. J.; Mimura, Y.; Dwek, R. A.; Rudd, P. M.; Elliott, T.; Davis, B. J. *Chem. Commun.* **2006**, 1401-1403.
81. Nakano, J.; Ichiyanagi, T.; Ohta, H.; Ito, Y. *Tetrahedron Lett.* **2003**, *44*, 2853-2856.
82. Tanaka, H.; Iwata, Y.; Takahashi, D.; Adachi, M.; Takahashi, T. *J. Am. Chem. Soc.* **2005**, *127*, 1630-1631.
83. Unverzagt, C. *Tetrahedron Lett.* **1997**, *38*, 5627-5630.
84. Hang, H. C.; Bertozzi, C. R. *Acc. Chem. Res.* **2001**, *34*, 727-736.
85. Macmillan, D.; Daines, A. M.; Bayrhuber, M.; Flitsch, S. L. *Org. Lett.* **2002**, *4*, 1467-1470.
86. Bengtsson, M.; Broddefalk, J.; Dahmen, J.; Hendrikson, K.; Kihlberg, J.; Lohn, H.; Srinivasa, B. R.; Stenvall, K. *Glycoconjugate J.* **1998**, *15*, 223-231.
87. Davis, B. J.; Lloyd, R. C.; Jones, J. B. *J. Org. Chem.* **1998**, *63*, 9614-9615.
88. Macindoe, W. M.; van Oijen, A. H.; Boons, G.-J. *Chem. Commun.* **1998**, 847-848.
89. Zhu, X.; Pachamuthu, K.; Schmidt, R. R. *J. Org. Chem.* **2003**, *68*, 5641-5651.
90. Zhu, X.; Haag, T.; Schmidt, R. R. *Org. Biomol. Chem.* **2004**, *2*, 31-33.
91. Thayer, D. A.; Yu, H. N.; Galan, M. C.; Wong, C.-H. *Angew. Chem. Int. Edn.* **2005**, *44*, 4596-4599.
92. Carrasco, M. R.; Brown, R. T. *J. Org. Chem.* **2003**, *68*, 8853-8858.
93. Peluso, S.; Imperiali, B. *Tetrahedron Lett.* **2001**, *42*, 2085-2087.
94. Marcaurelle, L. A.; Rodriguez, E. C.; Bertozzi, C. R. *Tetrahedron Lett.* **1998**, *39*, 8417-8420.
95. Rodriguez, E. C.; Marcaurelle, L. A.; Bertozzi, C. R. *J. Org. Chem.* **1998**, *63*, 7134-7135.
96. Huisgen, R. *Angew. Chem. Int. Edn.* **1963**, *2*, 565-598.
97. Marmuse, L.; Nepogodiev, S. A.; Field, R. A. *Org. Biomol. Chem.* **2005**, *3*, 2225-2227.
98. Joosten, J. A. F.; Tholen, N. T. H.; Ait El Maate, F.; Brouwer, A. J.; van Esse, G. W.; Rijkers, D. T. S.; Liskamp, R. M. J.; Pieters, R. J. *Eur. J. Org. Chem.* **2005**, 3182-3185.

99. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Edn.* **2002**, *41*, 2596-2599.
100. Kolb, C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Edn.* **2001**, *40*, 2004-2021.
101. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193.
102. Speers, A. E.; Cravatt, B. F. *Chem. Biol.* **2004**, *11*, 535-546.
103. Breinbauer, R.; Koehn, M. *ChemBioChem* **2003**, *4*, 1147-1149.
104. Macmillan, D.; Blanc, J. *Org. Biomol. Chem.* **2006**, *4*, 2847-2850.
105. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2004**, *126*, 15046-15047.
106. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776-779.
107. Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128*, 6640-6646.
108. Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684-11689.
109. Winans, K. A.; King, D. S.; Rao, V. R.; Bertozzi, C. R. *Biochemistry* **1999**, *38*, 11700-11710.
110. Bang, D.; Makhatadze, G. I.; Tereshko, V.; Kossiakoff, A. A.; Kent, S. B. *Angew. Chem. Int. Edn.* **2005**, *44*, 3852-3856.
111. Bang, D.; Kent, S. B. *Angew. Chem. Int. Edn.* **2004**, *43*, 2534-2538.
112. Wu, B.; Warren, J. D.; Chen, J.; Chen, G.; Hua, Z.; Danishefsky, S. J. *Tetrahedron Lett.* **2006**, *47*, 5219-5223.
113. Wu, B.; Chen, J.; Warren, J. D.; Chen, G.; Hua, Z.; Danishefsky, S. J. *Angew. Chem. Int. Edn.* **2006**, *45*, 4116-4125.
114. Bang, D.; Kent, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5014-5019.
115. Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705-6710.
116. Cotton, G. J.; Muir, T. W. *Chem. Biol.* **1999**, *6*, 247-256.
117. Hofmann, R. M.; Muir, T. W. *Curr. Opin. Biotechnol.* **2002**, *13*, 297-303.
118. Macmillan, D.; Bertozzi, C. R. *Angew. Chem. Int. Edn.* **2004**, *43*, 1355-1359.
119. Lue, R. Y. P.; Chen, G. Y. J.; Hu, Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2004**, *126*, 1055-1062.
120. Wood, R. J.; Pascoe, D. D.; Brown, Z. K.; Medlicott, E. M.; Kriek, M.; Neylon, C.; Roach, P. L. *Bioconjugate Chem.* **2004**, *15*, 366-372.
121. Lu, W.; Gong, D.; Bar-Sagi, D.; Cole, P. A. *Mol. Cell* **2001**, *8*, 759-769.
122. Tolbert, T. J.; Wong, C.-H. *Angew. Chem. Int. Edn.* **2002**, *41*, 2171-2174.
123. Macmillan, D.; Arham, L. *J. Am. Chem. Soc.* **2004**, *126*, 9530-9531.
124. Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123*, 526-533.
125. Guo, Z.-W.; Nakahara, Y.; Nakahara, Y.; Ogawa, T. *Carbohydr. Res.* **1997**, *303*, 373-377.
126. Hondal, R. J.; Nilsson, B. L.; Raines, R. T. *J. Am. Chem. Soc.* **2001**, *123*, 5140-5141.
127. Quaderer, R.; Hilvert, D. *Chem. Commun.* **2002**, 2620-2621.
128. Canne, L. E.; Bark, S. J.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 5891-5896.
129. Offer, J.; Boddy, C. N. C.; Dawson, P. E. *J. Am. Chem. Soc.* **2002**, *124*, 4642-4646.
130. Vizzavona, J.; Dick, F.; Vorherr, T. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1963-1965.
131. Tam, J. P.; Yu, Q. *Biopolym.* **1998**, *46*, 319-327.

132. Botti, P.; Carrasco, M. R.; Kent, S. B. H. *Tetrahedron Lett.* **2001**, *42*, 1831-1833.
133. Tchertchian, S.; Hartley, O.; Botti, P. *J. Org. Chem.* **2004**, *69*, 9208-9214.
134. Low, D. W.; Hill, M. G.; Carrasco, M. R.; Kent, S. B. H.; Botti, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6554-6559.
135. Marinzi, C.; Offer, J.; Longhi, R.; Dawson, P. E. *Bioorg. Med. Chem.* **2004**, *12*, 2749-2757.
136. Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007-2010.
137. Saxon, E.; Armstrong, J. D.; Bertozzi, C. R. *Org. Lett.* **2000**, *2*, 2141-2143.
138. Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2000**, *2*, 1939-1941.
139. Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. *Org. Lett.* **2001**, *3*, 9-12.
140. Lin, F. L.; Hoyt, H. M.; van Halbeek, H.; Bergman, R. G.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2005**, *127*, 2686-2695..
141. Liu, L.; Hong, Z.-Y.; Wong, C.-H. *ChemBioChem* **2006**, *7*, 429-432.
142. Merkx, R.; Rijkers, D. T. S.; Kemmink, J.; Liskamp, R. M. J. *Tetrahedron Lett.* **2003**, *44*, 4515-4518.
143. Nilsson, B. L.; Hondal, R. J.; Soellner, M. B.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 5268-5269.
144. Tolbert, T. T.; Wong, C.-H. *J. Am. Chem. Soc.* **2000**, *122*, 5421-5428.
145. Bode, J. W.; Fox, R. M.; Baucom, K. D. *Angew. Chem. Int. Edn.* **2006**, *45*, 1248-1252.
146. Ottenheijm, H. C. J.; Herscheid, J. D. M. *Chem. Rev.* **1986**, *86*, 697-707.
147. Tokuyama, H.; Kuboyama, T.; Fukuyama, T. *Org. Synth.* **2003**, *80*, 207-213.
148. Wasserman, H. H.; Ho, W.-B. *J. Org. Chem.* **1994**, *59*, 4364-4366.
149. Carrillo, N.; Davalos, E. A.; Russak, J. A.; Bode, J. W. *J. Am. Chem. Soc.* **2006**, *128*, 1452-1453.
150. Aimoto, S. *Biopolymers* **1999**, *51*, 247-265.
151. Hojo, H.; Matsumoto, Y.; Nakahara, Y.; Ito, E.; Suzuki, Y.; Suzuki, M.; Suzuki, A.; Nakahara, Y. *J. Am. Chem. Soc.* **2005**, *127*, 13720-13725
152. Ishiwata, A.; Ichiyanagi, T.; Takatani, M.; Ito, Y. *Tetrahedron Lett.* **2003**, *44*, 3187-3190.
153. Tachibana, Y.; Matsubara, N.; Nakajima, F.; Tsuda, T.; Tsuda, S.; Monde, K.; Nishimura, S.-I. *Tetrahedron* **2002**, *58*, 10213-10224.
154. Bordusa, F. *Chem. Rev.* **2002**, *102*, 4817-4868.
155. Zhong, Z.; Bibbs, J. A.; Yuan, W.; Wong, C.-H. *J. Am. Chem. Soc.* **1991**, *113*, 2259-2263.
156. Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. *Science* **1994**, *266*, 243-247.
157. Wong, C.-H.; Schuster, M.; Wang, P.; Sears, P. *J. Am. Chem. Soc.* **1993**, *115*, 5893-5901.
158. Witte, K.; Seitz, O.; Wong, C.-H. *J. Am. Chem. Soc.* **1998**, *120*, 1979-1989.
159. Canne, L. E.; Walker, S. M.; Kent, S. B. *Tetrahedron Lett.* **1995**, *36*, 1217-1220.
160. Li, X.; Kawakami, T.; Aimoto, S. *Tetrahedron Lett.* **1998**, *47*, 8669-8672.
161. Clippingdale, A. B.; Barrow, C. J.; Wade, J. D. *J. Pept. Sci.* **2000**, *6*, 225-234.
162. Swinnen, D.; Hilvert, D. *Org. Lett.* **2000**, *2*, 2439-2442.
163. Alsina, J.; Yokum, T. S.; Albericio, F.; Barany, G. *J. Org. Chem.* **1999**, *64*, 8761-8769.
164. Brask, J.; Albericio, F.; Jensen, K. J. *Org. Lett.* **2003**, *5*, 2951-2953.
165. Menzo, A. R.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 3885-3891.

166. Hackenberger, C. P. R.; Friel, C. T.; Radford, S. E.; Imperiali, B. *J. Am. Chem. Soc.* **2005**, *127*, 12882-12889.
167. von Eggelkraut-Gottanka, R.; Klose, A.; Beck-Sickinger, A. G.; Beyermann, M. *Tetrahedron Lett.* **2003**, *44*, 3551-3554.
168. Kenner, G. W.; McDermott, J. R.; Sheppard, R. C. *Chem. Commun.* **1971**, 636-637.
169. Heidler, P.; Link, A. *Bioorg. Med. Chem.* **2005**, *13*, 585-599.
170. Backes, B. J.; Ellman, J. A. *J. Org. Chem.* **1999**, *64*, 2322-2330.
171. Ingenito, R.; Drezнjak, D.; Guffler, S.; Wenschuh, H. *Org. Lett.* **2002**, *4*, 1187-1188.
172. Willoughby, C. A.; Hutchins, S. M.; Rosauer, K. G.; Dhar, M. J.; Chapman, K. T.; Chicchi, G. G.; Sadowski, S.; Weinberg, D. H.; Patel, S.; Malkowitz, L.; Di Salvo, J.; Pacholok, S. G.; Cheng, K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 93-96.
173. Ollivier, N.; Behr, J.-B.; El-Mahdi, O.; Blanpain, A.; Melnyk, O. *Org. Lett.* **2005**, *7*, 2647-2650.
174. Kawakami, T.; Sumida, M.; Nakamura, K.; Vorherr, T.; Aimoto, S. *Tetrahedron Lett.* **2005**, *46*, 8805-8807.
175. Chen, J.; Warren, J. D.; Wu, B.; Chen, G.; Wan, Q.; Danishefsky, S. J. *Tetrahedron Lett.* **2006**, *47*, 1969-1972.
176. Botti, P.; Villain, M.; Manganeillo, S.; Gaertner, H. *Org. Lett.* **2004**, *6*, 4861-4864.
177. Shekhar, S.; Ryberg, P.; Hartwig, J. F.; Mathew, J. S.; Blackmond, D. G.; Streiter, E. R.; Buchwald, S. L. *J. Am. Chem. Soc.* **2006**, *128*, 3584-3591.
178. Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **1997**, *62*, 1264-1267.
179. Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **2000**, *65*, 1144-1157.
180. Murata, M.; Buchwald, S. L. *Tetrahedron* **2004**, *60*, 7397-7403.
181. Hamann, B. C.; Hartwig, J. F. *J. Am. Chem. Soc.* **1998**, *120*, 7369-7370.
182. Mann, G.; Baranano, D.; Hartwig, J. F.; Rheingold, A. L.; Guzei, I. A. *J. Am. Chem. Soc.* **1998**, *120*, 9205-9219.
183. Frank, R. L.; Smith, P. V. *J. Am. Chem. Soc.* **1946**, *68*, 2103-2104.
184. Price, C. C.; Stacy, G. W. *J. Am. Chem. Soc.* **1946**, *68*, 498-500.
185. Molina, P.; Alajarin, M.; Vilaplana, M. J.; Katritzky, A. R. *Tetrahedron Lett.* **1985**, *26*, 469-472.
186. Brasdshaw, J. S.; Jones, B. A.; Gebhard, J. S. *J. Org. Chem.* **1983**, *48*, 1127-1129.
187. Zheng, N.; McWilliams, J. C.; Fleitz, F. J.; Armstrong, J. D.; Volante, R. P. *J. Org. Chem.***1998**, *63*, 9606-9607.
188. Li, G. Y.; Zheng, G.; Noonan, A. F. *J. Org. Chem.* **2001**, *66*, 8677-8681.
189. Cai, L.; Cuevas, J.; Peng, Y.-Y.; Pike, V. W. *Tetrahedron Lett.* **2006**, *47*, 4449-4452.
190. Kaiser, E.; Colescott, R. L.; Bossinger, C. D.; Cook, P. I. *Anal. Biochem.* **1970**, *34*, 595-598.
191. Matsueda, R.; Kimura, T.; Kaiser, E. T.; Matsueda, G. R. *Chem. Lett.* **1981**, *10*, 737-740.
192. Smith, A. B.; Savinov, S. N.; Manjappara, U. V.; Chaiken, I. M. *Org. Lett.* **2002**, *4*, 4041-4044.
193. Clive, D. J.; Hisaindee, S.; Coltart, D. M. *J. Org. Chem.* **2003**, *68*, 9247-9254.
194. O'Donnell, M. J.; Polt, R. L. *J. Org. Chem.* **1982**, *47*, 2663-2666.
195. Williams, R. E.; Ehrlich, P. P.; Zhai, W.; Hendrix, J. *J. Org. Chem.* **1987**, *52*, 2615 - 2617.
196. Macmillan, D.; Bertozzi, C. R. *Tetrahedron* **2000**, *56*, 9515-9525.
197. Offer, J.; Dawson, P. E. *Org. Lett.* **2000**, *2*, 23-26.

198. Shiozaki, M.; Ishida, N.; Hiraoka, T.; Maruyama, H. *Tetrahedron* **1984**, *40*, 1795-1802.
199. Bertho, J.-N.; Loffet, A.; Pinel, C.; Reuther, F.; Sennyey, G. *Tetrahedron Lett*. **1991**, *32*, 1303-1307.
200. Price, T. S.; Jones, L. M. *J. Chem. Soc.* **1909**, *95*, 1725-1729.

ORGANIC LETTERS

2004
Vol. 6, No. 25
4659−4662

Rapid Synthesis of Acyl Transfer Auxiliaries for Cysteine-Free Native Glycopeptide Ligation

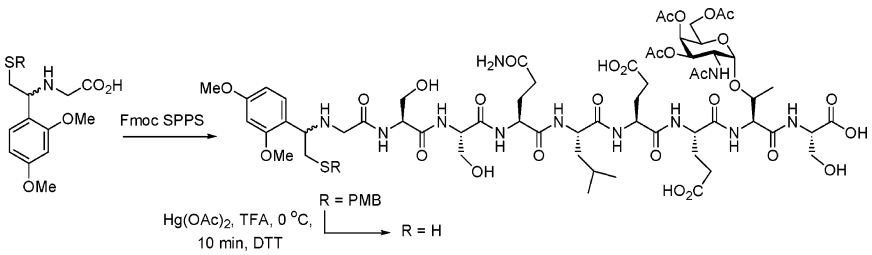
Derek Macmillan* and David W. Anderson

School of Chemistry, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh, EH9 3JJ, Scotland, UK

derek.macmillan@ed.ac.uk

Received September 14, 2004

ABSTRACT



Rapid, facile routes to the TFA-cleavable 4,5,6-trimethoxy-2-mercaptobenzyl and 1-(2,4-dimethoxyphenyl)-2-mercaptoethyl classes of auxiliaries for cysteine-free native chemical ligation are described. Rapid synthesis, coupled with mild cleavage conditions will undoubtedly broaden the utility of such auxiliaries, particularly where chemically fragile peptide modifications such as glycosylation are present.

The pioneering research of Kent[1] and Tam[2] in the development of chemoselective peptide ligation strategies has facilitated the total synthesis of hundreds of proteins.[3] However, the requirement for an N-terminal cysteine residue is considered a limitation of the native chemical ligation (NCL) methodology, and consequently cleavable thiol auxiliaries have been developed.[4] Most, however, required multistep syntheses and/or exposure to harsh conditions such as HF during synthesis or to facilitate auxiliary removal, which is not compatible with some posttranslational modifications such as glycosylation. We became interested in using the TFA−cleavable auxiliaries developed by Offer et al.[4f] and Botti et al.[4a] since, in our recent research, concerning the application of NCL to the assembly of O-linked glycoprotein GlyCAM-1, we had observed that a peptide containing both an N-terminal cysteine residue and an internal cysteine residue (introduced to facilitate a prior ligation) failed to give a ligation product until the internal cysteine was capped.[5] We then hoped to devise an alternative strategy employing a removable auxiliary to circumvent any problem arising from the presence of the internal thiol while, at the same time, developing an auxiliary synthesis that was more compatible with glycopeptide chemistry and that would allow convenient incorporation of auxiliaries as amino acid "cassettes" (Scheme 1) using standard Fmoc solid-phase peptide synthesis (SPPS). We focused on auxiliary-linked glycine conjugates since X-Gly[4d,g] ligation junctions seem most useful, particularly with the 1-phenyl-2-mercaptoethyl class of auxiliary,[4g] and these would adequately permit investigation of our preferred Leu-Gly and Ser-Gly ligation junctions.

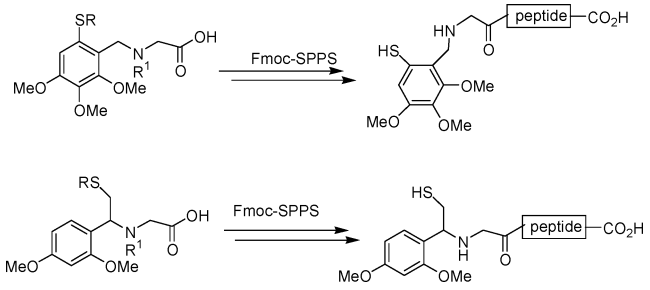
A key aim was to employ suitably labile protecting groups (R, Scheme 1) for the thiol functionality such as *S*-*p*-methoxybenzyl (PMB), *S*-*o*-nitrobenzyl (ONB), and *S*-trityl

(1) (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776.
(2) Liu, C.-F.; Tam, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6584.
(3) Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923.
(4) (a) Botti, P.; Carrasco, M. R.; Kent, S. B. H. *Tetrahedron Lett.* **2001**, *42*, 1831. (b) Canne, L. E.; Bark, S. J.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 5891. (c) Clive, D. L. J.; Hisaindee, S.; Coltart, D. M. *J. Org. Chem.* **2003**, *68*, 9247. (d) Low, D. W.; Hill, M. G.; Carrasco, M. R.; Kent, S. B. H.; Botti, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6554. (e) Offer, J.; Dawson, P. E. *Org. Lett.* **2000**, *2*, 23. (f) Offer, J.; Boddy, C. N. C.; Dawson, P. E. *J. Am. Chem. Soc.* **2002**, *124*, 4642. (g) Marinzi, C.; Offer, J.; Longhi, R.; Dawson, P. E. *Bioorg. Med. Chem* **2004**, *12*, 2749.

(5) Macmillan, D.; Bertozzi, C. R. *Angew. Chem., Int. Ed.* **2004**, *43*, 1355.

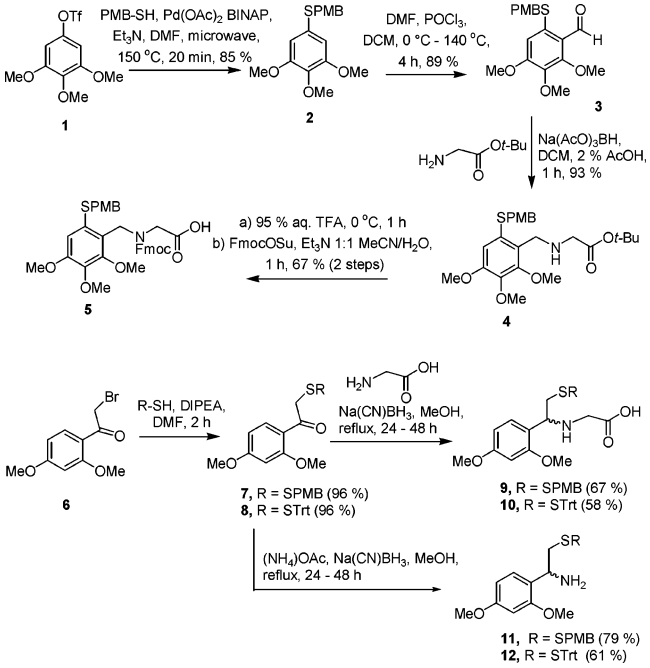
10.1021/ol048145o CCC: $27.50
Published on Web 11/17/2004

Scheme 1. Strategy for Introduction of Glycine-Linked 4,5,6-Trimethoxy-2-mercaptobenzyl and 1-(2,4-Dimethoxyphenyl)-2-mercaptoethyl Auxiliary Cassettes



(Trt) since reaction conditions employed for their removal have been shown to be compatible with glycopeptide synthesis.[4g,5,6] For the synthesis of glycine-linked 4,5,6-trimethoxy-2-mercaptobenzyl auxiliary **5** (Scheme 2), we

Scheme 2. Synthesis of Auxiliary-Linked Glycine Cassettes



introduced the SPMB-protected thiophenolic moiety using triflate **1** and PMB-thiol in a single step using the palladium-catalyzed coupling chemistry developed by Buchwald and Hartwig.[7]

Initially, we were disappointed to find that reported conditions for palladium-catalyzed aryl C−S bond formation using various substituted aryl triflates[8] failed to give appreciable quantities of thioether **2** with our substrate (**1**). However, this reaction occurred readily upon heating for 20 min at 150 °C in a microwave reactor. The commercially available aryl bromide could also be employed in place of **1**, but the yields were consistently lower (24−44%) under comparable reaction conditions. ONB-thiol failed to provide access to ONB-protected thiophenol under our optimized microwave conditions, and the *S*-trityl protecting group was not investigated at this stage as it was unlikely to survive the subsequent reaction conditions. Following Vilsmeier formylation,[4f] aldehyde **3** was conjugated directly to glycine using H-Gly-O*t*Bu (utilizing the *tert*-butyl ester to ease purification of the product) via reductive amination employing sodium triacetoxyborohydride as reductant in excellent yield.[9] The auxiliary−glycine conjugate **4** was then prepared for standard SPPS via TFA-mediated cleavage of the *tert*-butyl ester followed by Fmoc protection of the secondary amine. Synthesis of protected 1-(2,4-dimethoxyphenyl)-2-mercaptoethyl auxiliaries **9** and **10** was completed in only two steps. PMB- and trityl-protected thiol moieties were introduced by the action of the corresponding thiols on commercially available bromoacetophenone **6** to afford thioethers **7** and **8** in excellent yields. Glycine was then introduced via reductive amination using sodium cyanoborohydride as a reductant in refluxing methanol.[10] This route was investigated since under identical conditions for the preparation of **4** ($H_2NCH_2CO_2tBu$, $Na(AcO)_3BH$, AcOH/DCM), only reduction of the ketone was observed. Furthermore, only the glycine-linked auxiliary **4** was amenable to the *tert*-butyl deprotection and Fmoc protection cycle (**9** was resistant to Fmoc protection). We reasoned that, if the increased steric bulk proximal to the amino group of **9** was precluding Fmoc protection, then Fmoc protection might not be required at all during peptide synthesis. Consequently, **5** and **9** or **10** were introduced directly (using 3−5 equiv) to preassembled peptides corresponding to selected GlyCAM-1 protein fragments (Table 1) with no evidence of multiply coupled species arising from the use of **9** in SPPS. Interestingly, auxiliaries **11** and **12** of the type commonly employed in the submonomer approach (through subsequent reaction with bromoacylated peptides)[4] could also be conveniently prepared under identical reaction conditions employing ammonium acetate in place of glycine in the reductive amination. After Fmoc deprotection and cleavage from the resin of **5**-linked peptide **13** (Table 1), the PMB protecting group was efficiently removed using excess $Hg(OAc)_2$ in 10% aqueous AcOH followed by the addition of DTT to a final concentration of 5% w/v. **9**-linked peptides (**14** and **15**) were more resistant to such treatment, but the SPMB group was readily cleaved upon exposure to $Hg(OAc)_2$ in neat TFA for 10 min at 0 °C followed by dilution to 10% aqueous

(6) (a) Marcaurelle, L. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 1587. (b) Nicolaou, K. C.; Watanabe, N.; Li, J.; Pastor, J.; Winssinger, N. *Angew. Chem., Int. Ed.* **1998**, *37*, 1559.

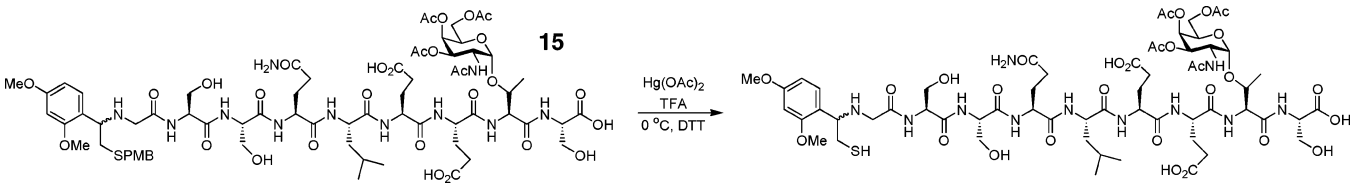
(7) (a)Murata, M.; Buchwald, S. L. *Tetrahedron* **2004**, *60*, 7397. (b) Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **2000**, *65*, 1144. (c) Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **1997**, *62*, 1264. (d) Marcoux, J.-F.; Wagaw, S.; Buchwald, S. L. *J. Org. Chem.* **1997**, *62*, 1568. (e) Mann, G.; Baranano, D.; Hartwig, J. F.; Rheingold, A. L.; Guzei, I. A. *J. Am. Chem. Soc.* **1998**, *120*, 9205.

(8) Zheng, N.; McWilliams, J. C.; Fleitz, F. J.; Armstrong, J. D., III; Volante, R. P. *J. Org. Chem.* **1998**, *63*, 9606.

(9) H_2N-Ala-O*t*Bu was also introduced under identical conditions in 96% yield.

(10) Williams, R. E.; Ehrlich, P. P.; Zhai, W.; Hendrix, J. *J. Org. Chem.* **1987**, *52*, 2615.

Table 1. Synthesis and Ligation Reactions of Auxiliary-Linked GlyCAM-1 Fragments[11]



		protected peptide−auxiliary conjugates		free thiol[a]			ligation product		
		m/z (calcd)	*m/z* (obsd)	*m/z* (calcd)	*m/z* (obsd)	thioester for model ligation[b]	*m/z* (calcd)	*m/z* (obsd)	auxiliary removed (Y/N)
13[c]	**5**-GlyCAM-1 (109−132)	2782.1	2782.2	2662.0	2661.7	AEEE**L**-SBn	3233.7	–[d]	N
						GLR**G**-SBn	3045.6	3046.0	Y
14	**9**-GlyCAM-1 (109−132)	2766.1	2766.4	2646.1	2645.0	AEEE**L**-SBn	2029.6	–[d]	N
15	**9**-GlyCAM-1 (75−83)	1582.6	1582.9	1462.5	1462.7	GLR**S**-SBn	1845.9	–[d]	N
						GlyCAM-1(1−77**S**)-SMESNA	9905.2	–[d]	N
						GLR**G**-SBn	1875.9	1876.6	Y
16[e]	**10**-GlyCAM-1 (75−83)	–	–	1462.5	1462.7	–	–	–	–

[a] After removal of SPMB or STrt protecting groups. [b] Prepared according to ref 12, and ref 13 for bacterially derived thioester. [c] Fmoc removed prior to cleavage from resin. [d] Only a trace amount of product was observed after 48 h by LC-MS and was not isolated. [e] TFA-mediated cleavage from the resin affords the fully unprotected auxiliary-linked glycopeptide, which is identical to deprotected **15** and was not tested independently in ligation reactions.

TFA and addition of DTT to a final concentration of 5% w/v (Figure 1).

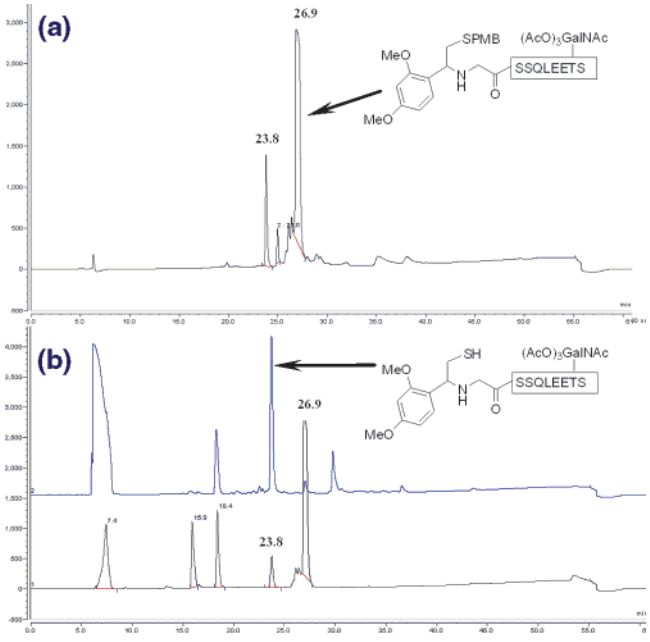


Figure 1. HPLC purification and SPMB deprotection of glycopeptide **15**. (a) TFA-cleaved **15** eluting at 26.9 min. Slight cleavage of SPMB under the resin cleavage conditions gives rise to fully unprotected material at 23.8 min; (b) SPMB cleavage from **15** in 10% AcOH, $Hg(OAc)_2$, 0 °C−rt, 5% DTT, 1 h (black trace) and in neat TFA at 0 °C, 10 min (blue trace) shows almost complete conversion.

With deprotected conjugates **13**−**15** in hand, we aimed to evaluate their ability to perform native chemical ligation reactions at our preferred Leu−Gly and Ser−Gly junctions, thus further probing the generality of the auxiliary-mediated ligation reaction with scaffolds arising from **5** and **9**. Model experiments were indeed required since there have been no published reports of the use of these particular scaffolds in cysteine-free NCL at Leu−Gly or Ser−Gly junctions.[4] For ligation reactions, AEEEL-SBn and GLRS-SBn thioesters were prepared so as to model the GlyCAM-1 Leu^{108}−Gly^{109} and Ser^{74}−Gly^{75} ligation junctions, respectively, using established methodology.[12]

Ligation reactions at peptide concentrations of approximately 3 mM were conducted in 6 M guanidine hydrochloride, 200 mM sodium phosphate buffer; pH 8.0, containing 2% w/v mercaptoethanesulfonic acid (MESNA) and 20 mM tris-carboxyethylphosphine (TCEP). The reactions were monitored by LC-MS. Initially, minor peaks attributable to the ligation products between AEEEL-SBn and **13** or **14** (PMB removed) were observed; however, these species did not accumulate over time and so we concluded that they were likely to be the transthioesterified yet unrearranged starting materials. A further ligation between fully deprotected glycopeptide **15** and GLRS-SBn thioester was also unsuccessful, as was ligation between unprotected **15** and a readily available bacterially derived peptide-thioester (corresponding to GlyCAM-1 residues 1-77, which also terminates in a serine thioester). Again only traces of product were observed after 48 h. This bacterially derived peptide-thioester had success-

(11) See Supporting Information for full details of peptide sequences, experimental procedures, and spectra.

(12) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684.

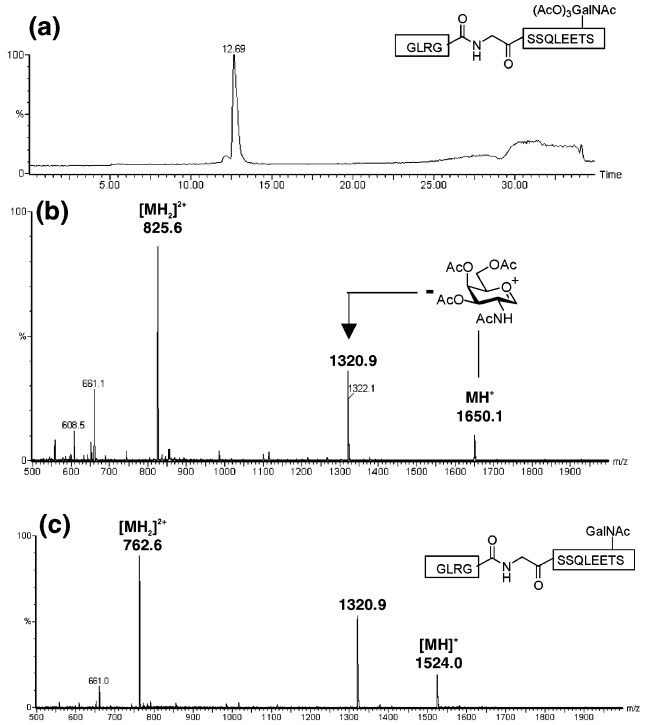


Figure 2. Cleavage of auxiliaries after Gly−Gly native glycopeptide ligation reactions. (a) LC-MS trace for purified ligated glycopeptide product after auxiliary removal. (b) MS of glycopeptide product (calcd mass = 1649.4 Da, showing characteristic fragmentation). (c) Treatment with hydrazine hydrate (2% v/v, 4 h) removes acetate esters on monosaccharide (calcd mass for deacetylated glycopeptide = 1524.4 Da).

fully participated in several NCL reactions previously with glycopeptides bearing a cysteine residue at the N-terminus.[5,13] These findings were in stark contrast to those obtained when a GLRG-SBn thioester (the glycine analogue of GLRS-SBn) was employed in ligation reactions. Reactions between GLRG-SBn thioester and unprotected **13** or **15** were virtually complete within 48 h.[11] The ligation products were readily purified by semipreparative reverse-phase HPLC, and the auxiliaries were removed upon treatment with 95% TFA for 3 h (Figure 2). The acetyl esters employed as protecting groups for the carbohydrate hydroxyl groups during SPPS were finally removed with 2% v/v hydrazine hydrate in 10 mM sodium phosphate buffer, which also confirmed the presence of a stable amide-linked ligation product.

In summary, we have developed particularly rapid and facile routes to two popular classes of TFA-cleavable acyl transfer auxiliaries for cysteine-free NCL. In both cases, deprotection of a *p*-methoxybenzyl thioether facilitated rapid liberation of the thiol functionality under conditions compatible with glycopeptide synthesis. Trityl-protected auxiliaries **10** and **12** could also be prepared in only two steps and incorporated into synthetic glycopeptides, though yields for the reductive amination of the trityl-protected precursors were consistently lower than those for the analogous SPMB-protected compounds. Additionally, we applied a relatively mild reductive amination protocol for the apparently problematic[4c] introduction of the amine functionality of auxiliaries **9**−**12**. Furthermore, unprotected glycine can participate in the reductive amination and the auxiliary−glycine conjugates obtained can be coupled directly to synthetic peptides without further protecting group manipulations. Auxiliary introduction and cleavage were shown to be compatible with the presence of glycosidic linkages and to function in cysteine-free NCL reactions across Gly−Gly ligation junctions. Unfortunately, both classes of auxiliary failed to deliver ligation products at Leu−Gly junctions and the 1-(2,4-dimethoxyphenyl)-2-mercaptoethyl auxiliary failed to deliver a ligation product at Ser−Gly junctions, which may limit their use in our studies unless amino acid substitutions in the GlyCAM-1 protein backbone are tolerated. We believe, however, that the rapid synthesis, coupled with mild cleavage conditions will undoubtedly broaden the utility of such auxiliaries in favorable cases, particularly where sensitive peptide modifications such as glycosylation are present.

Acknowledgment. The authors thank Dr. G. Tennant for helpful discussions and acknowledge financial support from The Royal Society, Cyclacel, and the BBSRC.

Supporting Information Available: Full synthetic procedures for the preparation of all compounds and data for characterization of all products. This material is available free of charge via the Internet at http://pubs.acs.org.

OL048145O

(13) Macmillan, D.; Bertozzi, C. R. *Tetrahedron* **2000**, *56*, 9515.